THIRD EYE ADVISORY

TECH mahindra

Whitepaper by
Tech Mahindra & Third Eye

Security by Design, Secure in Practice: An End-to-End Framework for GenAI

Authors
Sanjeev Mehrotra
Neil Kell

Executive Summary

As organizations rush to adopt generative AI (GenAI), security is transforming how organizations create content, write code, and automate workflows. Yet, as adoption surges, security often becomes a bolt-on, considered only after models move into production. This reactive approach leaves organizations vulnerable to unfamiliar threats, such as prompt injection, which undermines trust and slows innovation.

This paper makes a contrarian argument: security is not an obstacle but a catalyst for successful GenAI deployments. Drawing on practical experience, it demonstrates the need to embed security from day one, outlines common GenAI vulnerabilities, and presents an end-to-end framework for secure AI development, starting from the design phase.

It also highlights that by prioritizing security from the start, organizations can reduce risk, accelerate the adoption of GenAI, and build a sustainable competitive advantage.

Introduction: The GenAI Gold Rush and Hidden Risks

GenAI is revising the traditional cybersecurity playbook. There is immense pressure to implement large language models (LLMs) in different applications, from code generation to customer service. In the urgency to deliver fast, security is often treated as a last-minute compliance check rather than a foundational necessity. This requires a shift in mindset toward an end-to-end approach, where security is a continuous thread woven through design, development, deployment, and ongoing operation. The reality is that established security protocols aren't fully equipped to handle the novel risks of GenAI, which exist within AI's own logic and training data.

As a result, organizations face serious repercussions: sensitive data leaks through prompts, LLM outputs are manipulated by bad actors, and compliance audits become nightmares. Case in point, the security flaw in 2024 exposed thousands of private corporate GitHub repositories after Microsoft's AI tools accidentally cached and leaked sensitive data. Over 16,000 organizations were affected, as intellectual property from major tech companies, including access keys and security tokens, was compromised, thereby giving unauthorized users access to their systems.

As a result, enterprising teams must incorporate security directly into the design and development process from the beginning to transform GenAI from a liability into a reliable business enabler.

Common Blind Spots in GenAI Deployments

Vulnerability	Impact
Prompt injection	Manipulated outputs or disclosure of private data
Data leakage	Sensitive data exposed during training/inference
Shadow AI	Unauthorized GenAI apps outside IT governance
Model drift	Outputs degrade over time without monitoring
Lack of access controls	Broader risk of misuse and data exposure

Why Security by Design Matters

Technical vulnerabilities in GenAI cannot be treated as isolated IT problems, as they can cascade into serious business repercussions. These risks move beyond the technical realm, directly affecting organizations across three critical domains of finance, regulation, and stakeholder trust:

- **Financial Repercussions**
  Addressing vulnerabilities late in the GenAI development lifecycle leads to substantially higher monetary and operational costs. As systems mature, these flaws deeply embed in data pipelines and model architectures, making remediation complex and resource-intensive. It often involves untangling existing issues, reengineering workflows, and even retraining models that can cost up to five times more than proactive security integration. Designing security early enables seamless controls without disrupting innovation pipelines.

Differentiators That Drive Performance

Every feature, from omnichannel intake to real-time dashboards, plugs into your workflow without fuss. You get fault tolerance and local flexibility so operations keep running, even when things go wrong.

Feature	Why It Matters
Multi-Agent Architecture	Enables modular, scalable automation with fault tolerance.
Omnichannel Intake	Captures data from all relevant sources, reducing blind spots.
Regulatory Intelligence	Ensures timely and compliant reporting.
Translation & Transcription	Supports global operations and multilingual submissions.
Duplicate Detection	Cuts down on rework and improves data integrity.
Dashboards	Empowers teams with real-time insights and audit readiness.
Cloud Flexibility	Fits seamlessly into existing IT ecosystems.

- **Regulatory Exposure**
  Emerging AI regulations, such as the European Union (EU) AI Act and the Network and Information Security 2 (NIS2) Directive, require comprehensive risk assessments, transparent documentation, and robust technical and organizational safeguards. Non-compliance may result in severe financial penalties, reputational damage, and restrictions on AI solution deployment. Building security from the outset ensures organizations remain audit-ready and resilient to regulatory changes.

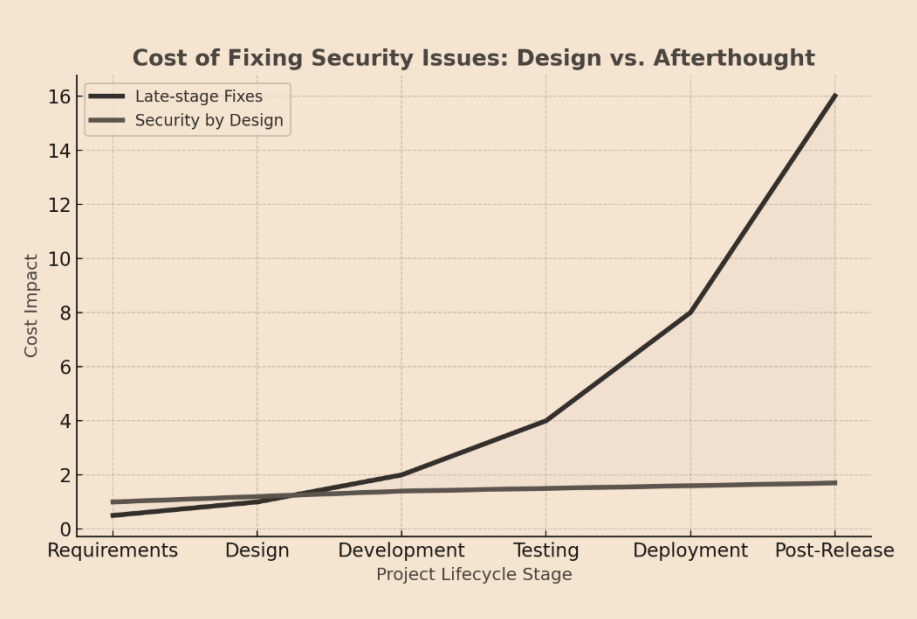
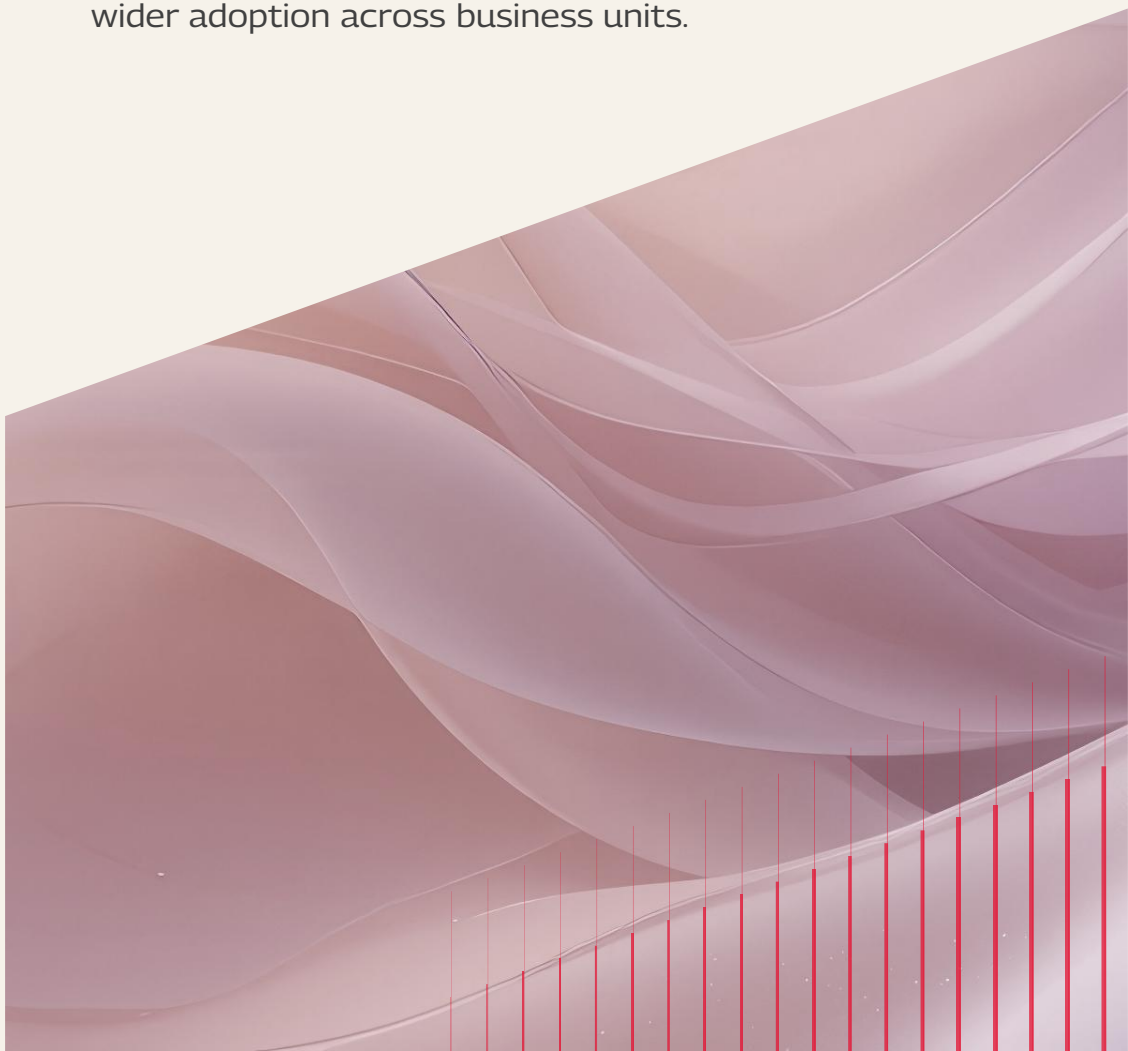
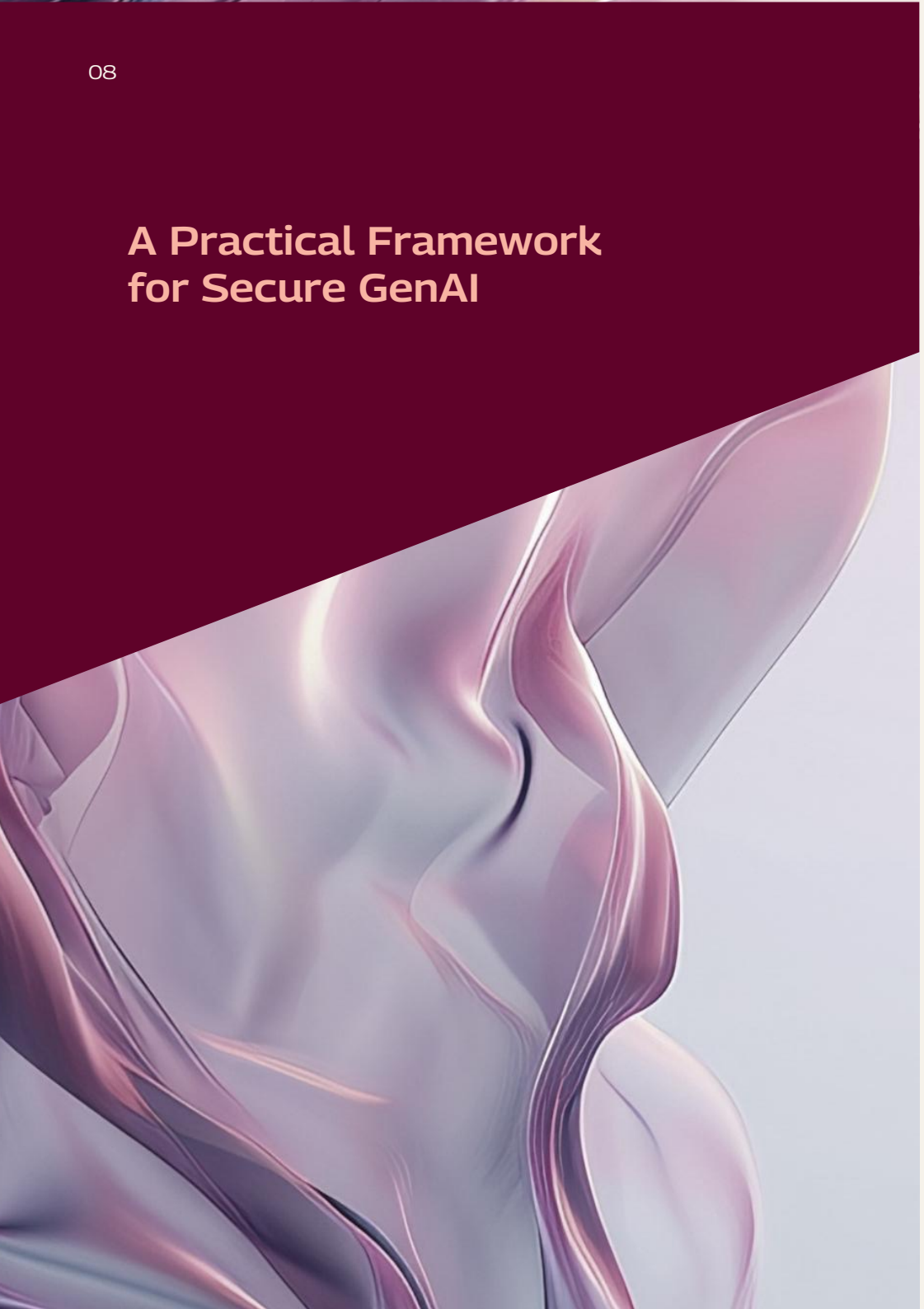


Figure 1: Cost of Fixing Security Issues: Design Vs Afterthought

- **Erosion of Trust**
  Data leakage, unauthorized access, or misuse of GenAI outputs can sabotage the confidence of stakeholders. A single breach can attract negative media attention, weaken customer relationships, and jeopardize partnerships. Embedding security at the design phase strengthens the reliability and credibility of AI initiatives, enabling confident and wider adoption across business units.



A Practical Framework for Secure GenAI



As organizations transition from experimentation to operational deployment of GenAI, they need to have a structured and defensible framework in place. According to a McKinsey flash survey of enterprise decision-makers, 63% of organizations rated GenAI implementation as a high or very high priority, yet 91% of these organizations did not feel very prepared to do so in a responsible way.[1] These gaps expose organizations to risks that scale rapidly as GenAI tools spread across departments.

A practical approach must be built on five core pillars:

- **Secure data sourcing**
  Before training any GenAI model, organizations must ensure that all data is rigorously anonymized and filtered for sensitive information. Employing data minimization, redaction, and encryption techniques protects personal and proprietary information, mitigating exposure risks during both training and inference.

- **Threat Modeling AI Pipelines**
  Conducting comprehensive threat assessments across the AI lifecycle is critical. It needs to include modeling risks from the AI supply chain, such as vulnerabilities in pre-trained models or third-party data sources, alongside internal threats like prompt injection. By mapping out all components and data flows from ingestion to deployment, teams can design targeted and effective mitigation strategies before an incident occurs.

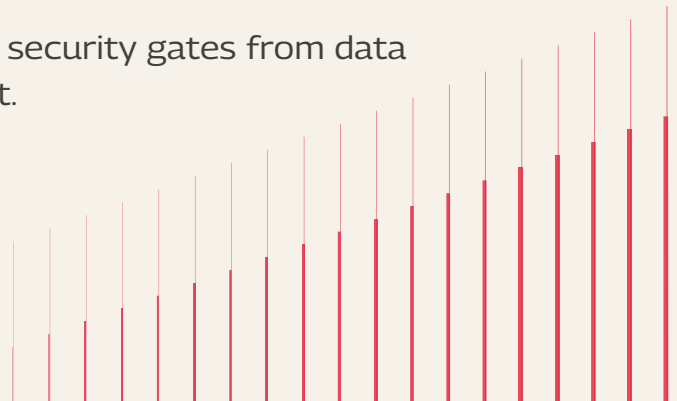
- **Identity and Access Controls**
  Implementing granular permissions ensures that only authorized users and systems can interact with GenAI models and their APIs. Applying principles of least privilege, multi-factor authentication, and robust role-based access management protects from misuse and limits the blast radius of any potential breach.

- **Continuous Monitoring**
  Maintaining real-time visibility into user inputs helps to prevent sensitive data leaks, such as personally identifiable information, thereby preserving both trust and compliance as systems scale. Moreover, ongoing surveillance of model behavior, data flows, and user prompts enables the early detection of anomalies such as model drift, malicious use, or policy violations. Real-time dashboards, alerting systems, and automated audits help to maintain operational integrity and respond to emerging threats.

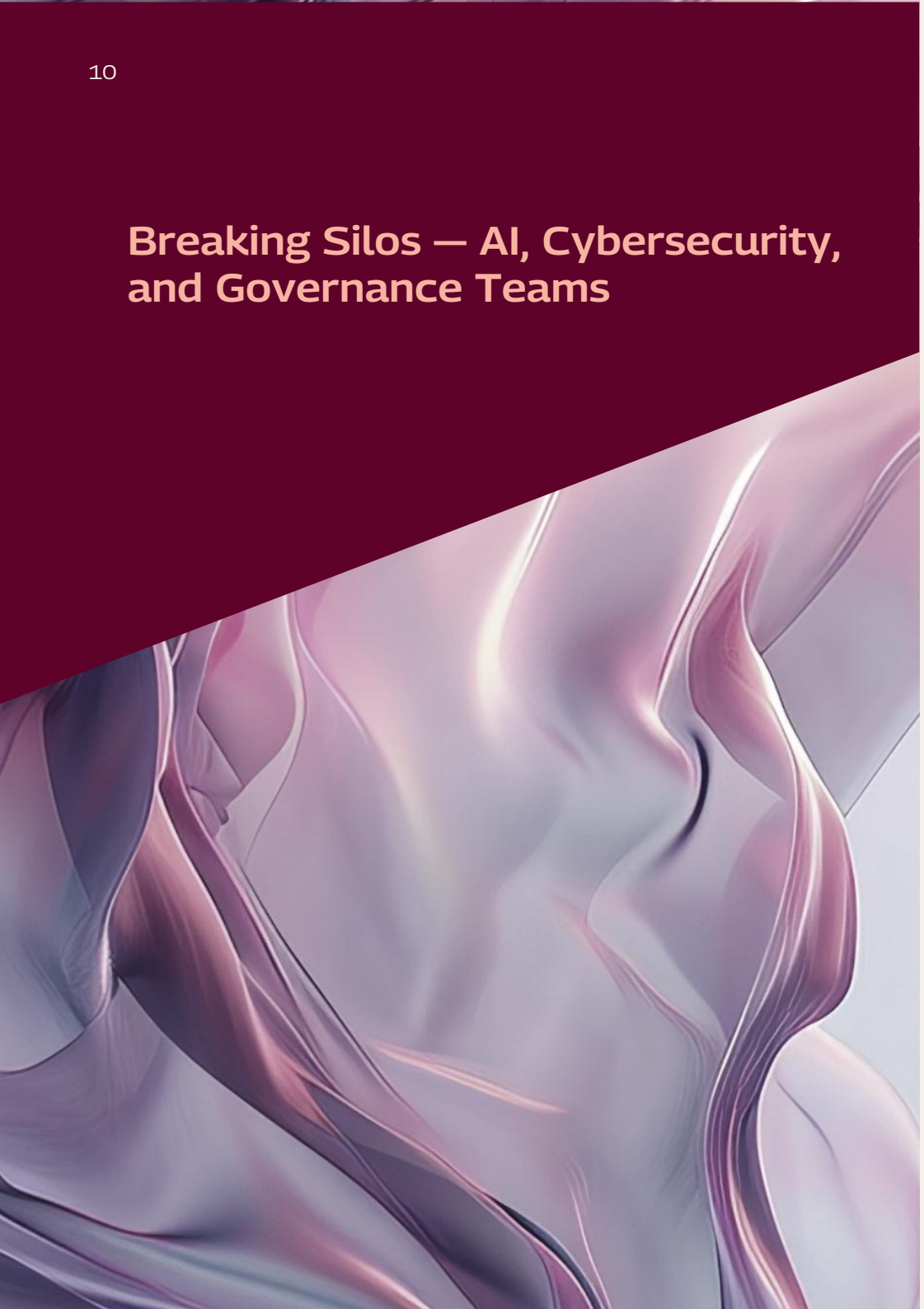
- **Explainability and Governance**
  Keeping comprehensive audit logs and making model decisions transparent are crucial for regulatory compliance and organizational trust. <ins>Effective governance also requires diligent documentation of the entire AI asset lifecycle. This includes recording the model's scope, limitations, data sources, retention periods, and review schedules. Such documentation is fundamental for supporting audits, proving compliance, and ensuring long-term transparency and accountability. Building explainability into GenAI systems allows stakeholders to understand model outputs, investigate decisions, and ensure accountability in AI-driven processes.

Visual idea

GenAI lifecycle showing security gates from data ingestion to deployment.



Breaking Silos — AI, Cybersecurity, and Governance Teams



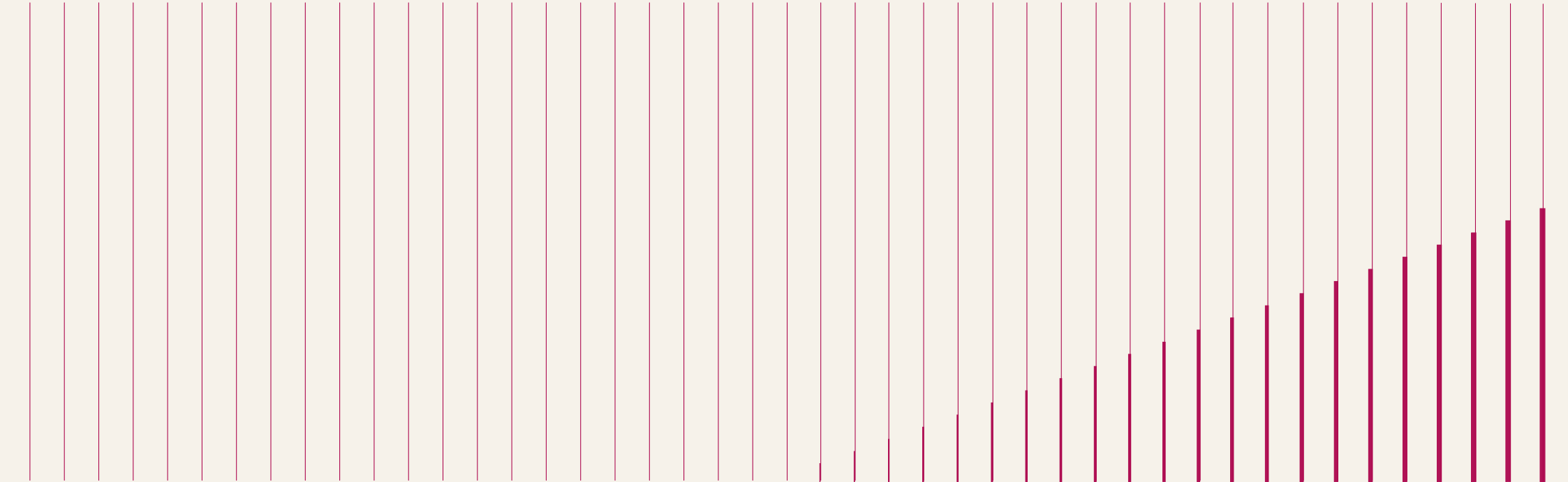
A practical framework is only effective if supported by an organizational culture that recognizes AI models and their underlying data as high-value corporate assets. Just like sensitive customer data or intellectual property, these AI assets must be protected with a commensurate level of security. This requires dismantling traditional silos between AI, cybersecurity, and governance teams to enable cross-disciplinary collaboration from day one. By integrating these functions early, organizations can transform security from a checkpoint into a catalyst for innovation across the enterprise.

To make this shift successful, they should focus on three core actions:

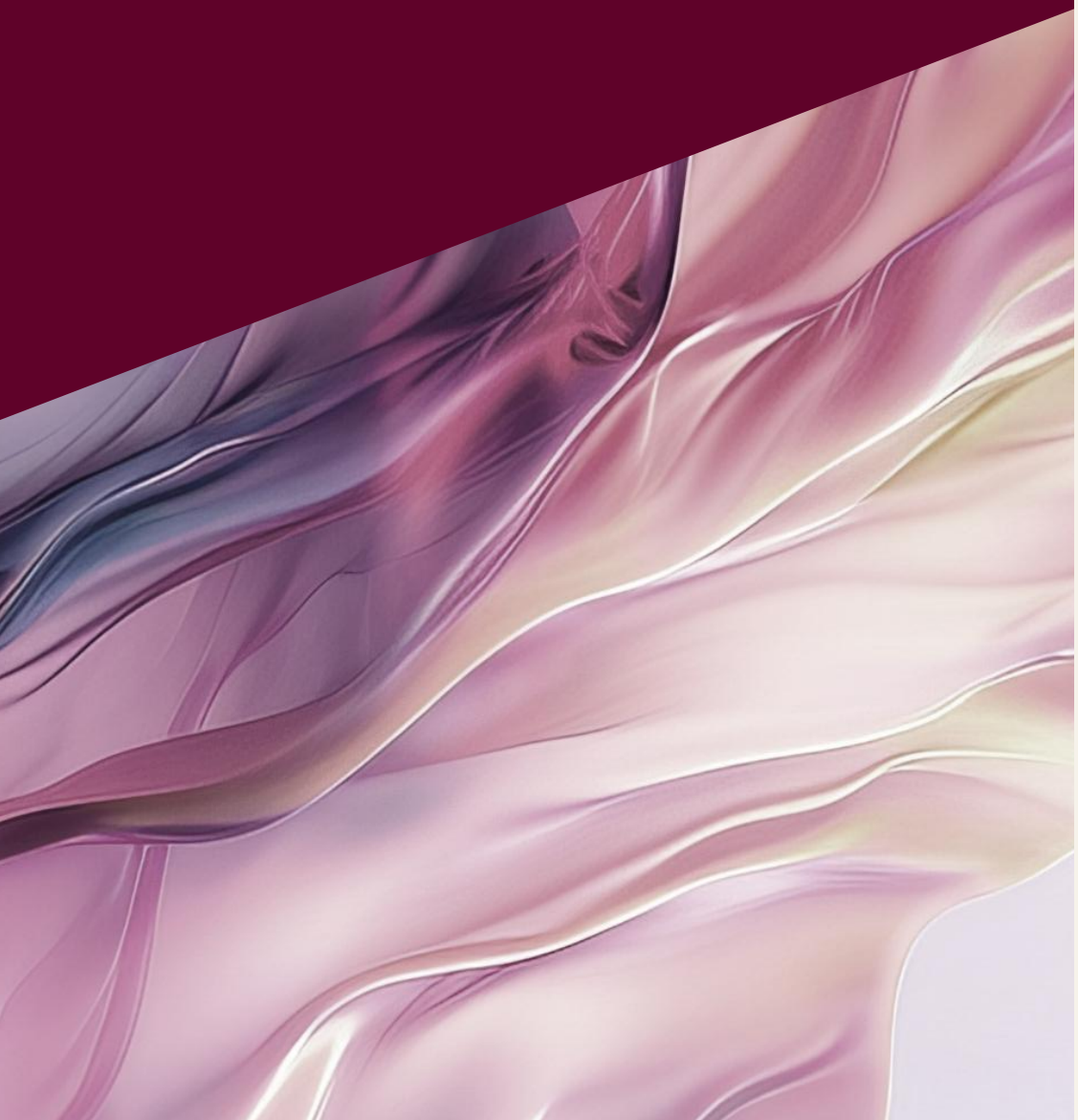
- **Integrate Security Experts from Day One**
  Onboarding cybersecurity professionals into AI project teams from inception is a critical first step. This must be paired with raising the security awareness of the AI and development staff themselves. This dual approach fosters a culture of shared responsibility for risk mitigation, ensuring that security is addressed proactively throughout the development lifecycle, rather than as a reactive afterthought.

- **Implement Shared Security Gates in MLOps Pipelines**
  Establishing security checkpoints throughout the machine learning operations (MLOps) pipeline, verify the integrity of data, models, and code at every (deployment) milestone. These gates facilitate automated vulnerability scanning, enforce policy compliance, and validate configurations before systems go live.

- **Establish Cross-Functional AI Risk Committees**
  Forming cross-functional committees that comprise stakeholders from IT, legal, compliance, and business units ensures comprehensive oversight and risk management. These bodies review GenAI deployments, assess risk profiles, and approve go-live decisions to enable faster, more coordinated responses to emerging threats.



Tech Mahindra's Approach



Tech Mahindra's approach to AI is defined by a commitment to secure, responsible, and resilient AI solutions where security controls are woven into the very fabric of every GenAI system we deliver. Rather than retrofitting protection after deployment, our teams architect platforms where data protection, governance, and regulatory alignment are foundational.
This proactive model both streamlines compliance and fortifies enterprises against evolving digital threats, supporting both internal operations and clients across diverse sectors. Our methodology is built on three core pillars:

- **Embedding Security Controls from the Ground Up**
  Our GenAI offerings, comprising TechM Orion, VerifAI, and our cybersecurity platform, are designed with security features intrinsic to their architecture. Core components include robust data encryption, rigorous access controls, audit logging, and continuous monitoring. These mechanisms are seamlessly integrated to ensure that data protection and governance are essential elements of the solution and not afterthoughts. For instance, our AI-powered security operations provide automated incident response, threat detection, and vulnerability prioritization, reducing manual remediation and improving compliance across the organization.

- **Explainability and Responsible Governance**
  Trust and transparency are central to Tech Mahindra's 'AI Delivered Right' philosophy. We leverage advanced explainability tools and frameworks, such as VerifAI, to validate AI outcomes across the discovery, development, and production phases. Our solutions visualize model reasoning and flag anomalies, and generate standardized, auditable reports. This level of transparency empowers organizations to demonstrate responsible AI practices, respond confidently to regulatory inquiries, and strengthen stakeholder trust. Throughout the AI lifecycle, our frameworks incorporate accountability, fairness, transparency, and explainability, thereby promoting the adoption of ethical AI.

- **Balancing Innovation and Compliance**
  Tech Mahindra's methodology enables clients to harness the full potential of GenAI while maintaining a steadfast commitment to compliance and risk governance. Our consulting and implementation practices harmonize rapid innovation with industry regulations and ethical frameworks, leveraging proprietary tools for risk assessment, compliance mapping, and responsible AI governance. This balanced approach accelerates time-to-market for transformative AI solutions, while ensuring risks are managed and regulatory standards are upheld.

Only 18% of security leaders say their organizations have an enterprise-wide responsible AI governance board.[2]

Conclusion: Security as an Innovation Accelerator

Positioning security as a foundational design principle lets organizations accelerate GenAI adoption, lower costs, and enhance stakeholder trust. Tech Mahindra, together with Third Eye, helps enterprises to implement this approach and create secure, scalable, and responsible GenAI systems that deliver lasting value.

By embedding security, explainability, and compliance from the outset, Tech Mahindra stands apart as a strategic partner for future-ready, reliable AI solutions. This comprehensive approach empowers enterprises to navigate the complexities of agentic technologies and the dynamic AI landscape with confidence and resilience.

Endnotes

1. McKinsey & Company. (2024, March 13). *Implementing generative AI with speed and safety*. McKinsey & Company. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/implementing-generative-ai-w...

2. McKinsey & Company. (2024, May 30). *The state of AI in early 2024: Gen AI adoption spikes and starts to generate value*. McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024

About the Author

Sanjeev Mehrotra is a dynamic technology leader with over 28 years of experience in driving digital security transformations for global enterprises. As the Global Head of Cybersecurity at Tech Mahindra, Sanjeev leads the vision, strategy, and execution of the company's cybersecurity portfolio, helping organizations build resilient, threat-aware environments in an increasingly complex digital world.

Prior to joining Tech Mahindra, Sanjeev spent over a decade and half at HCL, where he held key leadership positions and was instrumental in scaling the organization's global cybersecurity practice. His deep industry knowledge spans multiple sectors, including BFSI, healthcare, manufacturing, and telecom.

Sanjeev holds a management degree from the Institute of Management Technology (IMT) and a technical foundation built on years of hands-on experience in enterprise IT and security.

Based in Noida, India, Sanjeev plays a pivotal role in shaping Tech Mahindra's cybersecurity strategies, ensuring the delivery of cutting-edge solutions that help clients navigate the complex and ever-evolving digital landscape.

Sanjeev Mehrotra

Global Head - Cybersecurity,
Tech Mahindra

Analyst Detail

Neil Kell is one of the UK's leading security advisors. He operates at the strategic level in organisations, ensuring that key cyber security and resilience obligations are met. Neil has been a security professional for over 20 years. He was awarded the fellowship of the Institute of Consulting in recognition of his work and is one of a limited number of security professionals who hold the Lead Security and Information Risk Adviser (SIRA) designation under the UK National Cyber Security Centre's Certified Cyber Professional scheme. Outside of work Neil enjoys travel, rugby and has a passion for motoring. He is also a committed volunteer, occupying a prominent role in his local community.

Neil Kell

Expert Advisor Third Eye Advisory - Managed security services.

About **Tech Mahindra**

Tech Mahindra (NSE: TECHM) offers technology consulting and digital solutions to global enterprises across industries, enabling transformative scale at unparalleled speed. With 149,000+ professionals across 90+ countries helping 1100+ clients, Tech Mahindra provides a full spectrum of services including consulting, information technology, enterprise applications, business process services, engineering services, network services, customer experience & design, AI & analytics, and cloud & infrastructure services. It is the first Indian company in the world to have been awarded the Sustainable Markets Initiative's Terra Carta Seal, which recognizes global companies that are actively leading the charge to create a climate and nature-positive future. Tech Mahindra is part of the Mahindra Group, founded in 1945, one of the largest and most admired multinational federation of companies. For more information on how TechM can partner with you to meet your Scale at Speed™ imperatives, please visit https://www.techmahindra.com/.

www.techmahindra.com
www.linkedin.com/company/tech-mahindra
www.x.com/Tech_Mahindra