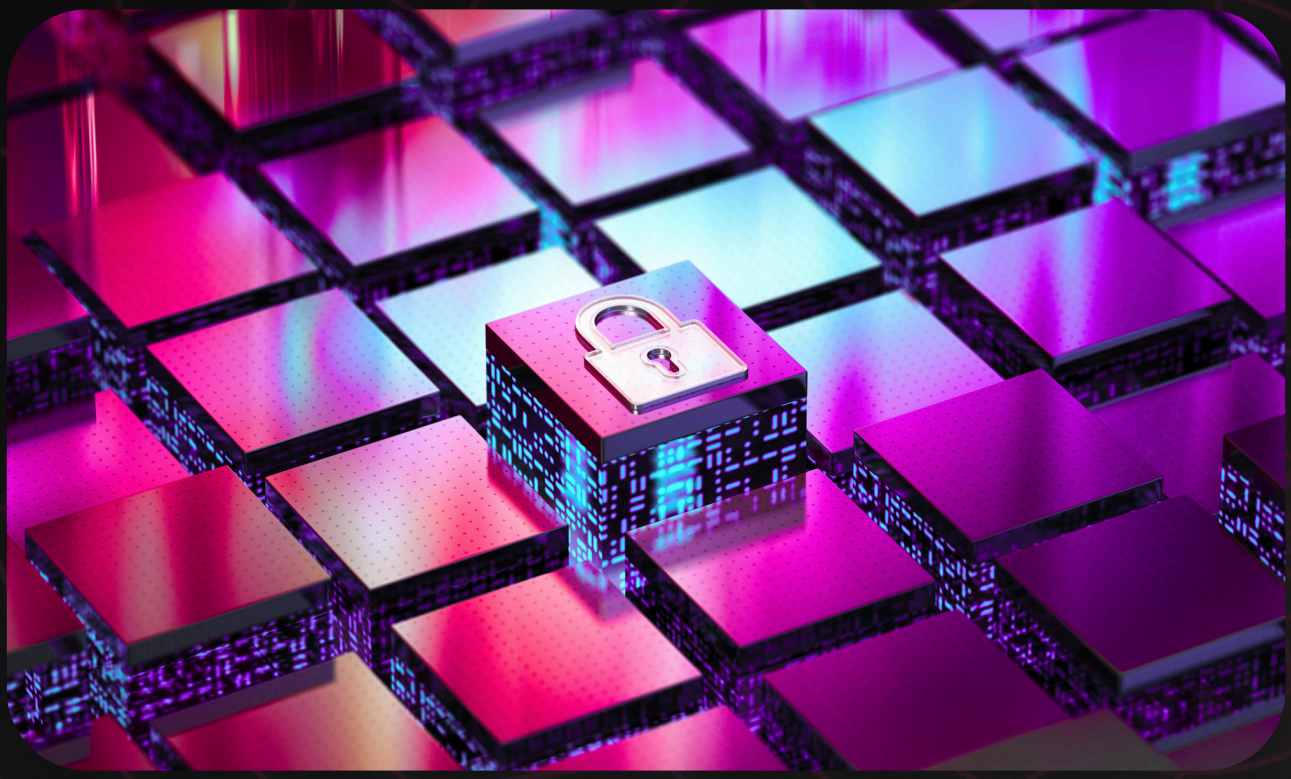
cyberkach

Low Confidence, High Risk

Benchmarking Organizational AI Security Readiness



DECEMBER 2025

Outline

Executive Summary

The speed of AI adoption has created a significant gap between technological capability and security maturity. Our analysis shows that enterprises are rushing to activate AI while largely failing to establish the foundational governance, controls, and ownership necessary to manage the resulting security risks. This creates a state of Low Confidence, High Risk.

50%

of security professionals lack confidence in their organization's ability to securely manage AI, confirming an unmanaged, high risk environment.
(Cyberkach Poll, 2025)

86% say their organizations have experienced AI-related security incidents in the past 12 months (Cisco, 2025).

Furthermore, the financial repercussions are severe: breaches involving unmanaged AI components, known as "Shadow AI," add an estimated $670,000 to the average cost of a data breach (IBM, 2025).

Immediate Action Points

Based on the severity of the readiness gap, security leadership must prioritize:

- **Mandate Clear Ownership**: Assign and empower a single, accountable executive (e.g., CISO) for AI risk and governance across all lines of business.
- **Close the Governance Gap**: Adopt and enforce an AI Governance Framework built on globally recognized standards (e.g., NIST AI RMF - Artificial Intelligence Risk Management Framework).
- **Invest in Defense**: Dedicate resources specifically to technical controls for securing the AI lifecycle, focusing on model integrity and prompt injection defense.

Methodology

The findings in this report are derived from a combination of proprietary, internal benchmarking data and rigorous analysis of major external industry reports, ensuring a comprehensive and corroborated view of the 2025 AI security landscape.

INTERNAL DATA COLLECTION

The internal data summarized in Section 4 (Cyberkach AI Security Readiness Benchmark) was collected via live polling during a three-day, cross-industry webinar series held in June and July 2025.

- **Audience**: Cybersecurity professionals, IT leaders, and business executives across various global industries.
- **Polling Dates**: July 2nd and July 9th, 2025 (Days 2 and 3 of the webinar series).
- **Data Focus**: Polling questions focused on internal AI governance policies, accountability structures, perceived barriers to readiness, and framework adoption status.

EXTERNAL DATA CORROBORATION

The internal findings were validated and enriched using data and statistics drawn from the following authoritative external sources including:

- Cisco 2025 Cybersecurity Readiness Index,
- Accenture “State of Cybersecurity Resilience 2025”
- Gartner 2025, Navigating Imminent AI Turbulence for Cybersecurity
- Netwrix 2025 Cybersecurity Trends global report
- IBM 2025 Cost of Data Breach Report
- SecurityScorecard 2025, Defending the Financial Supply Chain Report
- SecurityScorecard, 2025 Global Third-Party Breach Report.

Cross-Industry AI Security Trends

Major cross-industry AI security trends center on the shift to proactive, AI-driven defense mechanisms and the parallel challenge of combating AI-enhanced cyber threats. Organizations are widely adopting AI to automate threat detection, integrate predictive analytics, and improve overall operational efficiency across various sectors like finance, healthcare, and manufacturing

Accelerated AI deployment is outpacing control implementation, enabling threat actors to use AI for more sophisticated, large-scale attacks.

Defining the Terms
To understand the gap, it is crucial to define the core concepts driving risk:

AI DEPLOYMENT

AI Deployment is the process of integrating a validated AI model into a live, production environment where it can interact with users, data feeds, and business workflows to automate decisions or generate content.
60% *of organizations are already using AI in their IT infrastructure* (Netwrix, 2025).

AI SECURITY READINESS

AI security readiness is an organization's or country's preparedness to handle the security risks and challenges that come with implementing artificial intelligence. It involves assessing and improving capabilities across several areas, including data security and governance, model protection, workforce skills, and the integration of AI-specific security tools like AI firewalls.

Key Cross-Industry AI Security Trends

The rapid adoption of Artificial Intelligence is reshaping both defensive and offensive security strategies. The following trends highlight the core functional shifts and emerging risks dominating the 2025 security landscape.

Trend	Core Function	Defense/Impact
Predictive Threat Intelligence	Analyzing vast data sets using ML to anticipate TTPs.	Shifts defense from reactive blocking to proactive threat forecasting and vulnerability identification (e.g., CTI platforms like Cortex XSOAR).
Automated Detection and Response	Correlating data across domains (endpoint, cloud, identity) to automatically disrupt attacks.	Enables real-time threat containment and remediation, drastically reducing attack impact and analyst workload (e.g., XDR solutions).
Behavioral Analytics (UEBA)	Establishing dynamic "patterns of life" for users and systems.	Critical for detecting subtle insider threats or compromised accounts by flagging deviations from established norms (especially in Finance/Healthcare).
AI-Enhanced Detection	Using deep machine learning to spot sophisticated zero-day malware and adaptive phishing.	Bypasses traditional signature-based systems to prevent fileless and polymorphic malware before execution (e.g., Next-Gen AI Antivirus).
Cloud and IoT Security Automation	Continuous, 24/7 monitoring and anomaly detection across decentralized, multi-cloud and device environments.	Provides unified risk scoring and identifies high-risk misconfigurations across expansive, complex attack surfaces (e.g., CNAPPs like Wiz).
Adaptive Authentication & Zero Trust	Evaluating login parameters (IP, device, behavior) in real time to adjust authentication requirements.	Enforces a strong Zero Trust model by dynamically elevating trust or triggering MFA challenges based on real-time risk assessment.
Rise of AI-Powered Cybercrime	Attackers leveraging deepfakes, adaptive malware, and automated exploits.	Represents a significant counter-trend, forcing an "AI vs. AI" competition in defense, with deepfake social engineering leading to significant fraud (e.g., wire transfer scams).

Industry Specific AI Security Applications

The following table illustrates how the core, cross-industry AI security trends identified in the previous section are being strategically implemented.

Industry	Key AI Security Application	Primary Benefit
Finance	Real-time fraud detection and transaction monitoring	Reduced fraud rates and financial losses
Healthcare	Patient data protection and regulatory compliance (e.g., HIPAA)	Maintained patient trust and adherence to legal standards
Manufacturing	Predictive maintenance and securing industrial control systems	Enhanced operational resilience and protected production lines
Telecommunications	Network security, planning, and optimization	Improved operational security and network resilience

THE GAP BETWEEN AI SPEED AND AI SECURITY

External industry research reveals a clear and urgent pattern: the time required for attackers to exploit AI systems is shrinking, while the time required for enterprises to secure and control these systems remains largely unchanged. As organizations rush to deploy AI at scale, they unintentionally create a rapidly widening, unsecured perimeter that adversaries are already exploiting.

50% faster time to exploit account exposure through automated and AI-assisted techniques (Gartner, 2025).

Only **25%** of organizations have mature AI governance programs (AuditBoard, 2025).

Taken together, these trends show that enterprise AI deployments are scaling faster than their security controls, expanding an attack surface that is widely under protected.

Cyberkach AI Security Readiness Benchmark

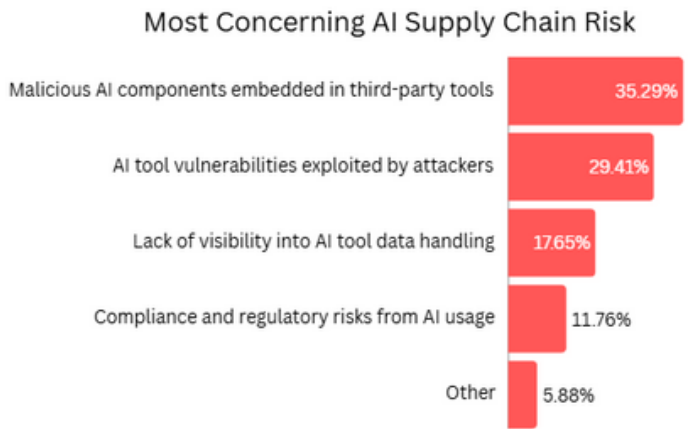
The internal Cyberkach benchmarking poll data reveals that security teams are currently working with insufficient support and fragmented authority, directly leading to a risk gap.

These insights are drawn from real-time polling conducted during Cyberkach's three-day, cross-industry webinar series held between June and July 2025, capturing the current state of enterprise AI security readiness.

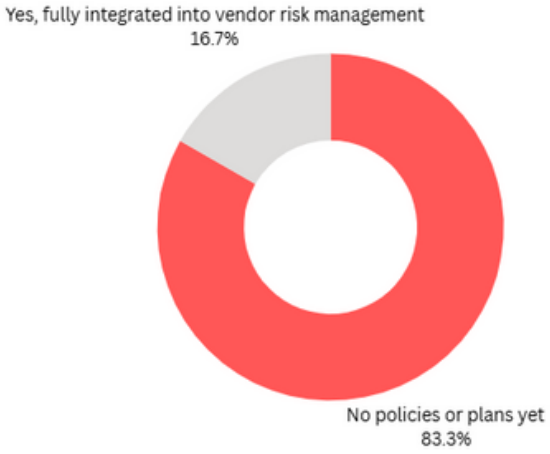
A. THE SUPPLY CHAIN CRISIS: HIGH FEAR, LOW VISIBILITY

Our polling identified the AI Supply Chain as the single greatest point of vulnerability. While respondents identified third-party risks as their top concern, their organizations are almost entirely unprepared to manage them.

- The Top Fear: When asked to identify the "Most Concerning AI Supply Chain Risk," 35.3% of respondents cited "Malicious AI components embedded in third-party tools." This fear outweighs compliance risks or simple visibility issues.
- The Reality Gap: Despite this fear, 83.3% of organizations reported having "No policies or plans yet" for third-party AI tool risk assessment.



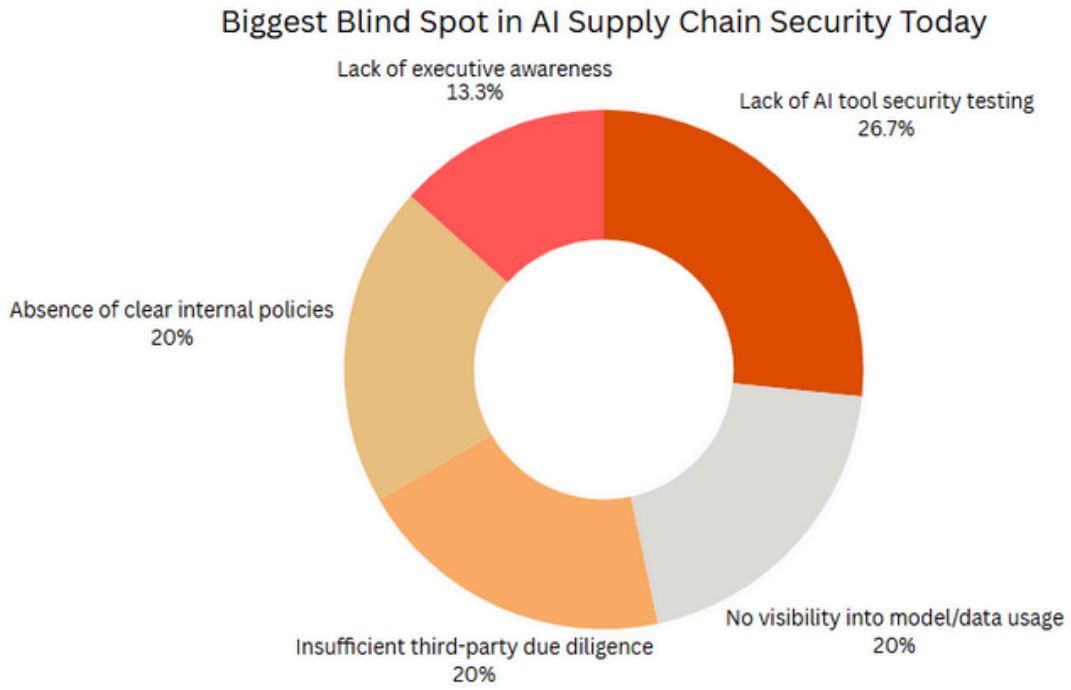
35.3% of respondents reported that "Malicious AI components embedded in third-party tools" is the Most Concerning AI Supply Chain Risk.



83.3% of respondents reported a lack of formal risk assessment procedures for third-party AI tools.

- The Blind Spot: The primary blind spot identified was a "Lack of AI tool security testing" (cited by **26.7%** of respondents), followed closely by a total lack of visibility into how vendors are handling data.

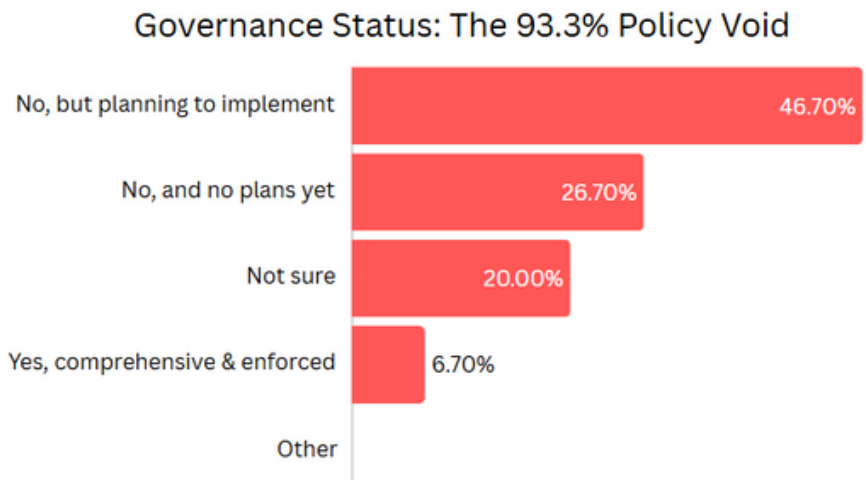
This internal data is supported by global threat intelligence. Data from Security Scorecard (2025) indicates that 98% of organizations have a relationship with at least one third-party vendor that has been breached.



Furthermore, in the Fintech sector specifically, **41.8%** of breaches are directly linked to third-party vendors. SecurityScorecard's research indicates that **35.5%** of all data breaches in 2024 were linked to third-party (supply chain) access. This is an increase from approximately 29% of breaches attributed to a third-party attack vector in 2023. The Cyberkach data confirms that this third-party gap is now extending into the AI landscape.

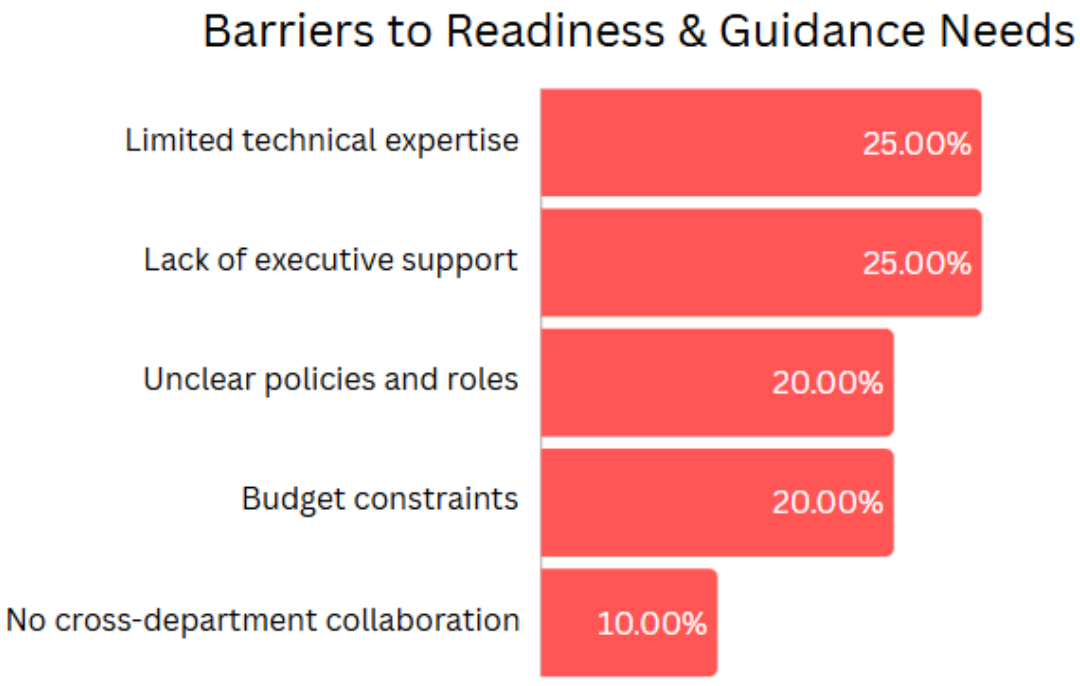
B. GOVERNANCE STATUS: THE 93.3% POLICY VOID

The **93%** governance gap and the lack of executive support/technical expertise are the primary internal barriers crippling organizational readiness.



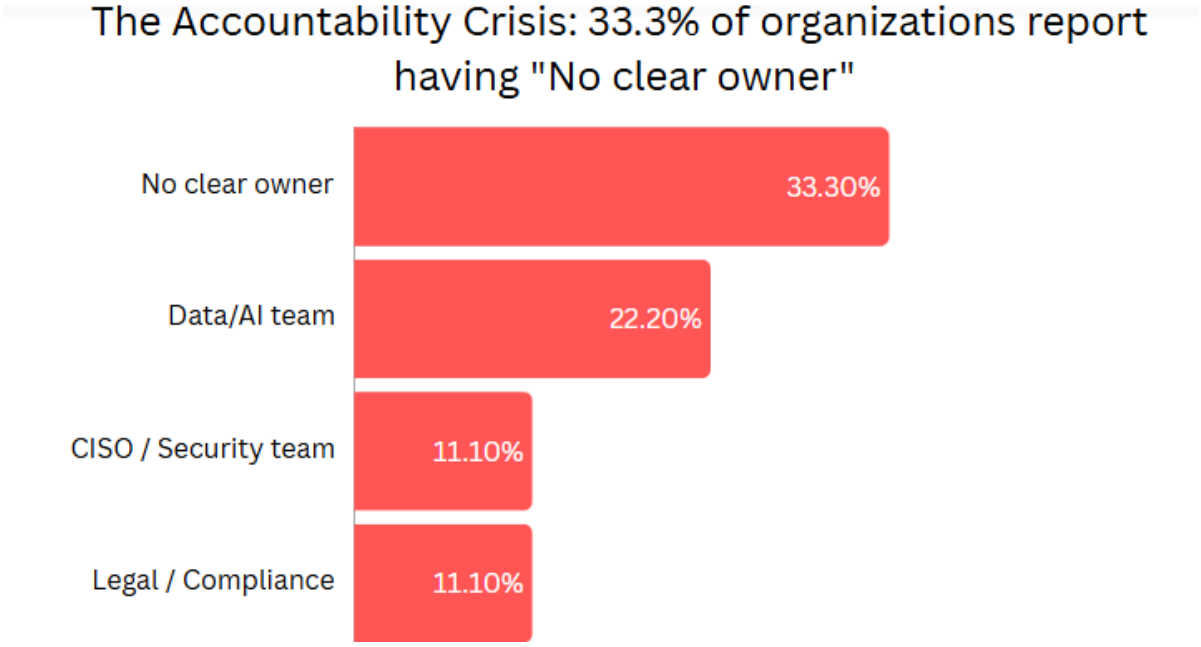
C. TOP OBSTACLES TO READINESS

When polled on the greatest challenges to achieving AI Security Readiness, the results overwhelmingly pointed to leadership and technical capability



D. THE ACCOUNTABILITY CRISIS

The security model fails when ownership is absent. **33.3%** of respondents report their organization having "No clear owner" for AI security governance, allowing risks to fester in the accountability void.



This comprehensive internal failure highlighted by the 83% governance gap and the accountability void means organizations are currently exposed to the most specialized and severe AI-native threats. The lack of foundational controls and dedicated expertise leaves organizations' AI infrastructure and implementation vulnerable to vector-specific attacks that bypass traditional network defenses.

AI Security Threat Vector Highlights

The most urgent, unmanaged threats are data and model integrity attacks (e.g., poisoning, evasion), which require specialized technical controls to mitigate. AI systems introduce unique attack surfaces that bypass conventional network and application defenses, creating blind spots for security teams.

AI Threat Vector	Description	Business Impact	Mitigation Focus
Model Poisoning	Injecting malicious, unlabeled data during the training phase to corrupt the model's accuracy and behavior.	Undermines the reliability and safety of the AI, leading to disastrous outcomes (e.g., misclassification of patients, erroneous financial fraud flags).	Data provenance, sanitization, and MLOps pipeline integrity checks.
Prompt Injection	Exploiting the LLM input mechanism to force the model to override its intended safety rules or expose proprietary training data.	Data leakage of confidential corporate information and business logic.	Output sanitization, content filtering, and technical controls.
Model Evasion (Adversarial ML)	Subtly manipulating inputs (e.g., adding noise to an image) to make the deployed model fail its classification task while remaining undetected by humans.	Allows attackers to bypass AI-based security controls (e.g., AI malware detection, deepfake identification).	Adversarial robustness testing and model monitoring.

The fact that 86% of organizations experiencing AI-related security incidents lacked proper AI access controls highlights the immediate need to secure the AI infrastructure itself.

Strategic Recommendations

Building AI Security Resilience

Security Leadership and the Executive Suite must take immediate action to mandate governance and ownership, formalize a framework, and invest in technical controls to move the organization out of the Low Confidence, High Risk zone. This strategy is built on three essential pillars below:

Strategic Governance and Policy

- **Establish a Formal Framework**: Implement a structured AI Security Governance model, specifically leveraging the NIST AI RMF (Artificial Intelligence Risk Management Framework) or similar standards, that mandates risk assessment before any new AI model is put into deployment.
- **Define Use Policies**: Create and socialize Acceptable Use Policies for both sanctioned (corporate LLMs) and unsanctioned ("Shadow") AI tools, clearly outlining data handling restrictions.
- **Update Vendor Risk Management**: Require explicit security and data integrity clauses for all third-party AI services, including cloud-hosted LLMs, APIs, and model providers.

Technical Controls & MLOps

- **Secure the AI Supply Chain**: Apply Zero Trust and strict access controls (MFA) to MLOps pipelines, training data, and model assets (Critical Infrastructure).
- **Test AI Attack Resistance**: Invest in tools to simulate and test models for Model Poisoning and Evasion attacks (Adversarial Robustness Testing) before deployment.
- **Continuous Runtime Monitoring**: Implement solutions to detect model drift and anomalous prompt patterns in live environments.

Accountability and Culture

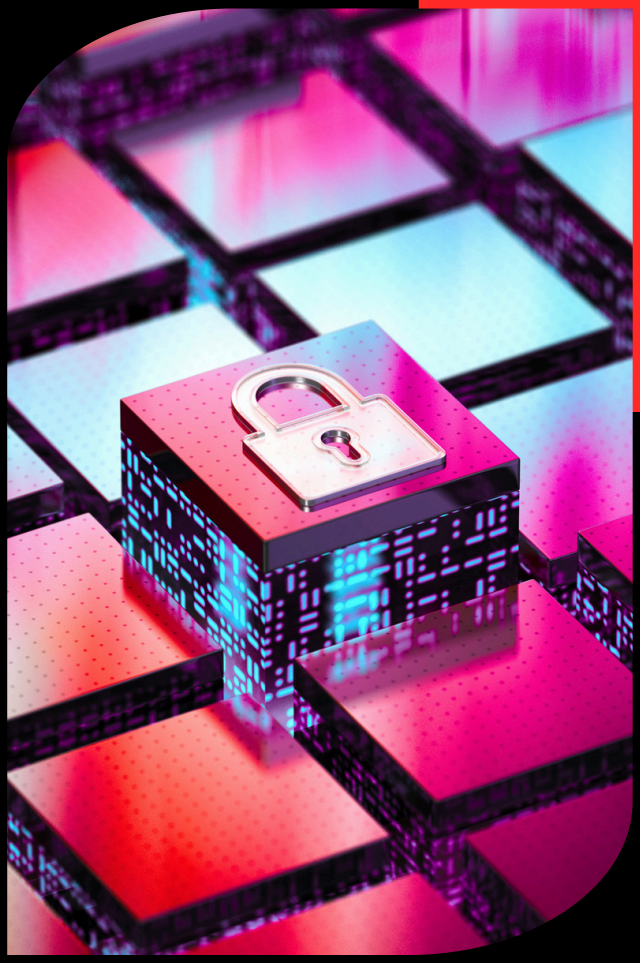
- **Executive Ownership**: The CISO or a delegated executive must be the single point of accountability for AI risk reporting to the Board.
- **Specialized Training:** Close the skills gap by funding advanced training for security engineers in AI Defenses (specifically how to counter attacks like data poisoning and evasion [what is commonly called Adversarial Machine Learning]) and secure MLOps practices (integrating security checks directly into the AI development and deployment pipeline).
- **Cultural Shift (Shift Left MLOps):** Promote cross-functional collaboration between Security and Data Science/MLOps teams to ensure security is "shifted left" into the earliest stages of the AI lifecycle. This means integrating security tools and processes (like adversarial robustness testing and data provenance checks) during the model training and development phase, rather than bolting them on post-deployment.

Sources

- Accenture, State of Cybersecurity Resilience 2025, https://www.accenture.com/us-en/insights/security/state-cybersecurity-2025
- AuditBoard, 2025 From blueprint to reality: Execute effective AI governance in a volatile landscape, https://auditboard.com/blog/new-research-finds-only-25-percent-of-organizations-report-a-fully-implemented-ai-governance-program
- Cisco, 2025 Cisco Cybersecurity Readiness Index, https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2025/m05/cybersecurity-readiness-index-2025.html.
- Cyberkach internal sources, 2025: Polling Data from Day 2 and 3 of Cyberkach AI Security Webinar June/July 2025, https://www.youtube.com/live/c-0VKz_cXZE?si=8z0B5I_ebhslMybE
- Gartner, "Predicts 2025: Navigating Imminent AI Turbulence for Cybersecurity", https://www.gartner.com/en/newsroom/press-releases/2025-03-18-gartner-predicts-ai-agents-will-reduce-the-time-it-takes-to-exploit-account-exposures-by-50-percent-by-2027
- IBM, 2025 Cost of Data Breach Report, https://newsroom.ibm.com/2025-07-30-ibm-report-13-of-organizations-reported-breaches-of-ai-models-or-applications,-97-of-which-reported-lacking-proper-ai-access-controls
- Netwrix, 2025 Cybersecurity Trends Report, https://netwrix.com/en/resources/research/2025-hybrid-security-trends-report/
- SecurityScorecard, 2025 Global Third-Party Breach Report Reveals Surge in Vendor-Driven Attacks, https://securityscorecard.com/company/press/securityscorecard-2025-global-third-party-breach-report-reveals-surge-in-vendor-driven-attacks/
- SecurityScorecard, 2025 Defending the Financial Supply Chain Report, https://securityscorecard.com/research/defending-the-financial-supply-chain/
- SecurityScorecard, 2023 Close Encounters of the Third (and Fourth) Party Kind, https://securityscorecard.com/research/cyentia-close-encounters-of-the-third-and-fourth-party-kind/

Contact Us

For further insights or expert interpretation, speak with our team. We help organizations translate threat intelligence into actionable insights. Reach out to us to continue the conversation and explore next steps.



Email
hello@cyberkach.com

Website
www.cyberkach.com

LinkedIn
Linkedin.com/company/cyberkach

Location
Lagos, NG