Flying Blind: The Enterprise AI Security Visibility Crisis

Why a Majority of CISOs Lack Visibility Into AI Deployments and What to Do About It

Unofficial AI-assisted Research

Cloud Security Alliance AI Safety Initiative

2026-03-18

Executive Summary

Enterprise AI adoption has accelerated substantially over the past three years, and the security infrastructure meant to govern it has not kept pace. A 2026 benchmark survey of 300 U.S. CISOs found that 67% report limited visibility into how AI is deployed across their organizations, and not a single respondent claimed full visibility [1]. This is not a marginal oversight challenge—it is a structural security problem with systemic causes and measurable consequences: shadow AI breaches cost organizations an average of $670,000 more than standard security incidents [2], and organizations with no AI-specific monitoring face threat exposure across model interfaces, data pipelines, agentic workflows, and third-party integrations that conventional tools were never designed to observe.

Three overlapping dynamics define this visibility deficit. First, employees and business units deploy AI tools independently, creating an estimated 1,200 unofficial AI applications per average enterprise [2]. Second, organizations are rapidly adopting agentic AI architectures—autonomous systems that chain decisions, invoke tools, and act on behalf of users—without the observability controls to understand what these agents are doing. Third, accountability for AI security remains fragmented across organizational functions, with no single owner responsible for maintaining a comprehensive view of AI deployments and their associated risks.

The consequences extend beyond data exposure. Security teams that cannot see AI are also unable to detect adversarial manipulation of AI behavior, unauthorized access to AI-connected systems, or data leakage through AI output channels. The attack surface is expanding faster than defenders can map it.

This whitepaper examines the root causes of the enterprise AI visibility crisis, quantifies its scope using the best available survey data, analyzes the threat landscape it creates, and maps a practical path forward grounded in CSA's AI Controls Matrix (AICM), the MAESTRO agentic AI threat modeling framework, Zero Trust architecture principles, and related CSA guidance. The central finding is that visibility is not merely a monitoring problem—it is a governance problem, and organizations that establish formal AI governance programs are significantly more likely to achieve the oversight capabilities that matter.

Introduction: The Scale of the Problem

The past three years have witnessed AI adoption in the enterprise transition from experimental initiative to operational reality with remarkable speed. Generative AI tools became embedded in daily workflows. Large language models were integrated into customer-facing applications, internal knowledge systems, and developer toolchains. And then, beginning in 2025, agentic AI—autonomous systems capable of planning, reasoning, and taking multi-step actions—began entering production environments at scale. According to CSA's 2025 State of AI Security and Governance survey, 19% of organizations were already using agentic AI, with 47% planning adoption within the following year [3]. A year earlier, a 2024 CSA survey of 2,486 IT and security professionals had already identified AI governance gaps as the sector's most pressing concern [12]—a finding that the 2025 and 2026 data show intensified rather than improved.

The pace of this adoption has created a structural imbalance. Security programs are designed around the principle of visibility: you cannot protect what you cannot see. But enterprise AI adoption has followed a pattern where deployment decisions are made by developers, business units, and individual employees far faster than security teams can inventory, assess, and monitor the resulting systems. The result is that a majority of organizations now operate AI systems they cannot fully observe, across threat surfaces they cannot fully characterize, with security controls that were designed for a fundamentally different era of enterprise technology.

CSA and others have begun using the term "AI visibility deficit" to describe this gap—a label that captures both the monitoring and governance dimensions of the challenge—and it is shaping up to be one of the defining security concerns of 2026. The stakes are high. AI systems touch sensitive data, make consequential decisions, interact with core business systems like ERP and CRM platforms, and increasingly take autonomous actions that affect customers, employees, and partners. Without visibility into these systems, organizations cannot manage the risks they pose.

The CSA AI Safety Initiative published this whitepaper to provide security leaders with a rigorous analysis of the visibility deficit, its structural causes, and a governance-grounded framework for closing it. The analysis draws on the best available survey data, CSA's own research corpus on AI security governance, and the emerging body of evidence on agentic AI security risks. The findings are both sobering and actionable: the visibility deficit is severe, but organizations that invest in governance-first approaches demonstrate measurably better security outcomes.

The Visibility Gap: Quantifying What CISOs Cannot See

The Core Finding

The scope of the enterprise AI visibility problem became sharply defined in early 2026 when Pentera published its AI Security and Exposure Benchmark Report, based on responses from 300 U.S. CISOs and senior security leaders. The headline finding—67% of CISOs report limited visibility into how AI is deployed across their organizations—was startling not simply for the number but for what accompanied it: not a single respondent reported full visibility [1]. In an environment where organizations are deploying AI tools at enterprise scale, the security function is largely operating without a complete picture.

The survey found that the top three barriers to securing AI infrastructure are lack of internal expertise (50%), limited visibility into AI usage (48%), and insufficient AI-specific security tools (36%) [1]. Notably, budget was not the primary constraint—only 17% cited budget limitations as their leading challenge. This suggests the visibility deficit is not primarily a resource problem. It is a knowledge, tooling, and governance problem: organizations lack the expertise to know what to look for, the tools to look for it, and the governance structures to ensure someone is responsible for looking.

The Monitoring Gap

The visibility deficit extends into the monitoring controls organizations have deployed for the AI systems they do know about. A separate analysis of agentic AI deployments found that only 38% of organizations monitor AI agent decisions end-to-end—tracking the full chain from prompt to tool invocation to output [16]. Only 41% have runtime guardrails in place to constrain agent behavior, and just 17% continuously monitor agent-to-agent interactions in multi-agent systems [16]. Perhaps most starkly, one in five organizations acknowledged deploying agents with no guardrails or monitoring whatsoever [16].

These are not edge cases. As agentic architectures move from pilot to production, the absence of monitoring becomes an absence of accountability. When an agent makes a decision that causes harm—accessing unauthorized data, executing an unintended action, leaking information through its outputs—an organization without end-to-end monitoring has few reliable means to detect the incident promptly, reconstruct the full decision chain, or prevent recurrence through targeted remediation.

The CSA 2025 State of AI Security and Governance survey corroborates this picture at the governance level: 73% of respondents reported being neutral or not confident in their organization's ability to execute an AI security strategy [3]. Only 26% reported having comprehensive AI security governance

policies in place. The gap between what is deployed and what is governed is significant—and growing.

The Tools Mismatch

A critical dimension of the visibility deficit is the mismatch between the tools organizations use to monitor AI and the tools required to monitor it effectively. The Pentera 2026 survey found that 75% of enterprise CISOs rely on legacy security controls—endpoint, application, cloud, or API security tools—to protect AI systems [1]. Only 11% of respondents reported having security tools specifically designed for AI systems [1].

This is not merely a technology gap—it reflects a category error in how AI risk is being managed. Endpoint security tools are designed to observe processes and file activity; they cannot interpret AI agent behavior. Network monitoring tools can observe API calls but are generally not calibrated to distinguish legitimate AI automation from a compromised agent—the behavioral signatures are different, and the detection logic has not been updated to account for them. Identity systems can track OAuth grants but are not calibrated to flag the accumulation of excessive permissions by an AI service account. The result is that organizations believe they are monitoring AI when they are, in practice, monitoring the infrastructure beneath AI—a fundamentally different signal.

Shadow AI: The Unsanctioned Deployment Problem

Scope and Prevalence

Shadow IT has been a persistent enterprise security challenge for decades; shadow AI represents an accelerated and more consequential version of the same dynamic. Research from 2025 and 2026 suggests that unsanctioned AI deployment is not a marginal phenomenon but a structural feature of how AI spreads through organizations. According to Microsoft and LinkedIn's Work Trend Index, 78% of AI users bring their own tools to work [15], and the average enterprise has approximately 1,200 unofficial AI applications in active use [2]. When employees adopt AI tools outside of IT and security oversight, they create deployment footprints that are, by definition, invisible to the security function.

The data exposure dimensions of shadow AI are severe. Research published in 2026 found that 63% of employees who used AI tools during 2025 pasted sensitive company data—including source code and customer records—into personal chatbot accounts [2]. Fifty-two percent of those employees were

reluctant to disclose this usage to their organizations [5]. This combination of widespread data exposure and deliberate concealment creates conditions where sensitive corporate and customer data is routinely flowing into AI systems that have not been assessed, approved, or monitored by security teams.

Financial Consequences

The financial consequences of shadow AI are measurable and significant. Shadow AI breaches cost organizations an average of $670,000 more than comparable incidents [2], and one in five organizations reported a breach linked to unauthorized AI use [2]. Among large enterprises—those with annual revenue above $1 billion—64% reported losing more than $1 million to AI failures [2]. These figures reflect a concrete and growing risk. Shadow AI is not an abstract governance concern; it is a demonstrated vector for expensive security incidents, and the majority of enterprises are currently operating with substantial exposure across this surface.

Why Traditional Approaches Fail

Shadow AI is particularly resistant to conventional security controls because it exploits the same channels that legitimate work traffic uses. Employees accessing a consumer AI service through a corporate browser generate web traffic indistinguishable from other browser activity without specialized monitoring. Data pasted into a chat interface traverses standard HTTPS channels without triggering data loss prevention rules that were designed for file transfers and email. The informal nature of shadow AI adoption means it leaves no procurement trail, no vendor contract, and no security assessment that might otherwise trigger review.

The CSA publication *AI Gone Wild: Why Shadow AI Is Your IT Team's Worst Nightmare* (March 2025) documented this challenge systematically, observing that shadow AI represents a convergence of employee productivity incentives with organizational governance failures [6]. The solution is not principally technical—it is a governance challenge that requires clear policies, accessible alternatives, and organizational cultures that treat AI security as a shared responsibility rather than a security team problem.

Agentic AI: The Emerging Blind Spot

The Nature of the Agentic Visibility Problem

The emergence of agentic AI systems—those capable of autonomous planning, tool use, multi-step reasoning, and action—represents a qualitative change in the visibility challenge. Conventional AI deployments are relatively static: a model receives a prompt and returns a response, and the exchange is discrete and bounded. Agentic systems are dynamic: they chain actions over time, invoke external tools and APIs, store and retrieve information across sessions, and may coordinate with other agents in complex pipelines. Each of these capabilities creates observability challenges that are poorly addressed by existing monitoring infrastructure.

A 2026 analysis found that only 21% of executives reported complete visibility into agent permissions, tool usage, or data access patterns—meaning that in 79% of organizations, deployed agents are taking actions, accessing data, and invoking tools in ways that security teams cannot fully observe [2]. Eighty percent of surveyed organizations reported having observed risky agent behaviors, including unauthorized system access and improper data exposure [2], and yet only 29% of organizations reported being prepared to secure agentic AI deployments [4].

This gap—between observed risk and security preparedness—is particularly acute because agentic systems often accumulate capabilities over time that exceed their original design scope. Agents frequently acquire long-lived credentials that may not be subject to standard rotation schedules, accumulate permissions through successive tool integrations, and operate across system boundaries in ways that are difficult to attribute to specific users or business functions. The identity and access management implications are significant and largely unaddressed.

The Tool Invocation Problem

When an AI agent invokes an external tool—executing a database query, calling an API, writing to a file system, or triggering a downstream action—the resulting event is often invisible to conventional security monitoring. Seventy-one percent of organizations report that AI tools now have access to core business systems including ERP and CRM platforms [16], but monitoring the boundary between AI agent behavior and those systems requires observability at the agent layer, not just at the infrastructure layer. Organizations that monitor at the infrastructure layer alone see the API call but cannot determine whether it was authorized, intended, or consistent with the agent's approved scope of action.

This creates a class of security events that existing tools cannot detect: an agent that has been manipulated through prompt injection, for instance, may execute actions that are technically valid from an access control perspective but that represent an adversarial takeover of the agent's decision-making. Without monitoring that captures the full decision chain—the prompt, the reasoning, the tool call, and the output—these incidents are invisible.

Multi-Agent Architectures Amplify the Risk

The challenge compounds significantly in multi-agent systems, where individual agents coordinate with one another to accomplish complex tasks. In these architectures, the decision chain may span multiple agent boundaries, and each handoff represents a point where monitoring continuity can break. Research found that only 17% of organizations continuously monitor agent-to-agent interactions [16], leaving the coordination layer of multi-agent systems almost entirely unobserved.

The security implications extend to adversarial scenarios. Research reported in 2026 found that multi-turn adversarial attacks combining jailbreak techniques achieved success rates as high as 92% across tested models, exploiting memory and tool access capabilities that single-turn protections failed to address [4]. In a multi-agent environment without end-to-end monitoring, these attacks could propagate across agent boundaries before any anomaly is detected.

Structural Causes of the Visibility Deficit

The Governance Fragmentation Problem

The visibility deficit is not primarily a technology problem. The CSA 2025 State of AI Security and Governance survey identified a structural cause that explains much of the monitoring gap: accountability for AI security is fragmented across multiple organizational functions, with no single owner responsible for maintaining a comprehensive view of AI deployments and their risks [3]. Fifty-three percent of respondents reported that the security team is primarily responsible for securing AI systems, but AI deployment decisions are made by a much more dispersed group: 20% of respondents identified a dedicated AI/ML team as the primary deployment function, 19% identified IT, 16% identified cross-functional groups, and additional responsibility was distributed across business unit owners, data science teams, and senior leadership [3]. The CISO's role has expanded correspondingly, with AI security governance decisions increasingly requiring cross-functional authority that extends well beyond a traditional security remit [14].

When deployment authority is distributed and security responsibility is nominally centralized in a team that has limited visibility into deployment decisions, the result is predictable: security teams learn about AI deployments after the fact, often through security incidents rather than governance processes. The CSA publication *AI Organizational Responsibilities: Governance, Risk Management, Compliance and Cultural Aspects* (2024) documents this dynamic and provides a framework for establishing clear ownership and accountability structures that connect deployment authority to security responsibility [7].

The Tooling Lag

Enterprise security tooling evolves in response to observed threats, and AI-specific threats are still relatively new in the industry's experience. The current generation of SIEM, EDR, CASB, and API security tools was designed and deployed to address cloud, SaaS, and endpoint risks that predate the current generation of AI architectures. Adapting these tools to monitor AI-specific behaviors—prompt flows, model output patterns, agent tool invocations, multi-model data pipelines—requires either significant customization of existing platforms or investment in purpose-built AI security tooling that the market is still maturing.

The 75% of CISOs who report relying on legacy controls for AI security are not making an irrational choice—in most cases, purpose-built alternatives do not yet exist at enterprise scale, or the investment case has not been established with sufficient urgency. But the consequence is a monitoring blind spot that grows as AI systems grow more complex and more consequential.

The Expertise Gap

The shortage of personnel with AI security expertise is a compounding factor that the data consistently surfaces. The Pentera 2026 survey found that lack of internal expertise was the single most-cited barrier to securing AI, reported by 50% of respondents [1]. The CSA 2025 survey found that 61% of organizations cite understanding AI risks as their top implementation hurdle, and 53% cite skills gaps [3].

This expertise deficit has a direct effect on visibility: organizations that do not understand AI-specific risks are poorly positioned to know what to monitor for, which signals matter, and how to interpret anomalies when they appear. Monitoring without expertise produces noise, not insight. The expertise gap is therefore not only a workforce problem—it creates a downstream visibility problem by degrading the quality of monitoring even where monitoring infrastructure exists.

The Governance Dividend

Governance as a Visibility Multiplier

One of the clearest findings in the available data is that formal AI governance is not merely a compliance function—it is the most reliable predictor of AI security maturity across all measurable dimensions, including visibility. The CSA 2025 State of AI Security and Governance survey found that organizations with comprehensive AI governance policies are twice as likely to adopt agentic AI with appropriate controls, three times more likely to train staff on AI security, and twice as confident in their ability to protect AI systems—a consistent pattern across governance metrics, though the findings are drawn from a 300-respondent survey [3]. CSA has characterized this pattern as a governance "maturity multiplier"—the mechanism by which formal governance programs convert AI investment into measurable security capability across the organization [13]. On the specific question of shadow AI, the survey found that governance programs were the primary mechanism organizations used to identify and manage unsanctioned deployments.

The mechanism is straightforward. Governance programs establish AI registries that require disclosure of AI deployments. They create approved-tool lists that channel employees toward sanctioned alternatives. They define accountability structures that ensure someone is responsible for knowing what AI is deployed and how it is secured. They establish monitoring requirements that make AI systems observable as a condition of deployment approval. Without these governance foundations, monitoring is reactive and incomplete—organizations learn about AI deployments when something goes wrong, not before.

The Cost of the Governance Gap

The obverse of the governance dividend is the governance gap cost. Organizations without formal AI governance have no mechanism to systematically discover shadow AI deployments, no process to assess AI systems before they go into production, and no accountability structure to ensure that security controls are maintained over time. The 74% of organizations that do not have comprehensive AI governance policies [3] are operating AI systems with correspondingly reduced visibility—and the financial consequences, as documented above, are measurable.

The governance gap also has a compounding quality. As AI systems proliferate, the cost of establishing governance after the fact increases. Each unsanctioned deployment creates dependencies that complicate remediation. Each unmonitored system becomes a potential pivot point for future incidents.

By analogy with cloud and IT governance experience, early investment in AI governance—before deployments scale—is likely to be significantly more efficient than remediation after the fact.

Building AI Security Visibility: A Framework-Driven Approach

Establishing the AI Asset Inventory

Visibility begins with inventory. Organizations cannot monitor AI systems they do not know about, and the starting point for addressing the visibility deficit is a systematic effort to discover and document all AI systems in use across the enterprise. This effort should encompass not only formally approved deployments but also the shadow AI landscape—the tool usage patterns that exist outside of IT approval processes.

An effective AI asset inventory program combines multiple discovery mechanisms: network traffic analysis to identify connections to AI APIs and services, procurement data review to identify AI-related vendor relationships, employee surveys and departmental interviews to surface informal tool usage, and integration with IT service management processes to capture AI projects as they are initiated. The CSA AI Controls Matrix (AICM) v1.0 provides a structured framework for categorizing AI assets across 18 security domains, including governance, supply chain, data management, and model operations [8]. Using AICM as the organizational backbone for the inventory ensures that asset records capture not only what AI systems exist but how they are positioned within the broader security architecture.

AICM's Shared Security Responsibility Model (SSRM) is particularly useful for clarifying which security controls are the responsibility of the organization versus the AI service provider, a distinction that is frequently unclear in practice and that directly affects what organizations need to monitor themselves versus what they can rely on providers to address [8].

Implementing AI-Appropriate Monitoring

Once an inventory exists, organizations need monitoring controls calibrated to AI-specific risks. This means extending beyond infrastructure-layer monitoring—which captures network traffic, API calls, and system events—to include agent-layer monitoring that tracks decision chains, tool invocations, and output patterns. Concretely, AI-appropriate monitoring should encompass prompt and completion

logging for all AI interactions with sensitive systems, tool call auditing for agentic systems that invoke external APIs or services, output scanning for anomalous data disclosure patterns, and agent-to-agent interaction monitoring in multi-agent architectures.

The Zero Trust principle of "never trust, always verify" maps cleanly to the AI monitoring challenge, because AI requests are inherently difficult to authenticate by origin or intent. In a Zero Trust architecture, every AI request, every tool invocation, and every data access by an AI system is treated as an untrusted action that requires explicit verification against policy. This approach, detailed in CSA's Zero Trust guidance and the *Using Zero Trust to Secure Enterprise Information in LLM Environments* publication [9], ensures that monitoring is continuous rather than event-driven, and that anomalies are detected through policy comparison rather than signature matching.

For agentic systems specifically, organizations should implement what security researchers have termed "agent accountability"—the continuous capture of the reasoning chain that connects prompts to actions, enabling forensic reconstruction of agent behavior after the fact. The MAESTRO framework for agentic AI threat modeling provides threat categorizations across seven layers of an AI agent's architecture, from the model layer through orchestration, memory, and tool access layers, and maps visibility requirements to each layer [10]. Using MAESTRO as a monitoring design guide ensures that surveillance is comprehensive across the full attack surface that agentic systems present.

Governance Structures for Sustainable Visibility

Monitoring infrastructure is necessary but not sufficient. For visibility to be sustained over time—as AI systems evolve, new deployments are introduced, and organizational contexts change—governance structures must ensure that visibility is maintained as a continuous organizational capability rather than a periodic assessment.

Effective AI governance for visibility purposes requires three components. First, a mandatory AI registration process that requires all AI deployments to be disclosed before going into production, enabling proactive inventory maintenance rather than reactive discovery. Second, a structured AI security assessment process, aligned with AICM control domains, that evaluates new deployments against defined security requirements before approval. Third, ongoing monitoring requirements tied to deployment approval—AI systems that cannot be monitored appropriately should not be approved for deployment, and AI systems that lose their monitoring coverage should be flagged for remediation.

The CSA publication *AI Organizational Responsibilities: Governance, Risk Management, Compliance and Cultural Aspects* provides detailed guidance on establishing these governance structures, including accountability model design, policy development, and cultural change management [7]. Its framework

for shadow AI prevention is particularly relevant to the visibility deficit: it recommends a combination of policy clarity, accessible sanctioned alternatives, and detection capability as the three levers most effective at reducing unsanctioned AI deployment.

CSA Resource Alignment

The visibility deficit described in this whitepaper connects to several bodies of guidance that CSA has developed to support organizations in managing AI security risks. Organizations addressing the challenges documented here can engage these resources in sequence, beginning with the governance and control framework and moving toward technical implementation and threat modeling.

The foundational resource is the **AI Controls Matrix (AICM) v1.0**, which establishes the governance and control framework for enterprise AI security across 18 security domains and 243 control objectives [8]. For the visibility problem specifically, AICM's governance, transparency, and monitoring domains define the controls organizations need to establish and maintain AI asset inventories, implement appropriate monitoring, and maintain audit trails across AI operations. Its Shared Security Responsibility Model (SSRM) is especially valuable here, clarifying which controls belong to the organization versus the AI service provider—a distinction that directly determines what organizations must monitor themselves versus what they can delegate to vendors. Working in concert with AICM, the **STAR for AI Registry** enables organizations to document their AI security posture using the AI Consensus Assessment Initiative Questionnaire (AI-CAIQ), providing a structured self-assessment mechanism that can serve as the foundation for an AI asset inventory and governance baseline [8]. Third parties can use STAR for AI to assess vendor security posture, which addresses the supply-chain dimension of the visibility problem.

For agentic AI deployments, **MAESTRO (Multi-Agent Environment Security Threat Review and Operations)** provides a threat modeling framework specifically designed for the attack surfaces that autonomous AI systems introduce [10]. Its seven-layer architecture maps vulnerabilities across model behavior, orchestration, memory management, tool access, and inter-agent trust—each layer representing a distinct monitoring requirement that conventional security tooling is not designed to address. Organizations should use MAESTRO both to design monitoring controls for existing agentic deployments and to inform incident response planning for the adversarial scenarios those deployments may face.

The access control and monitoring principles that underpin AI visibility are addressed in CSA's **Zero Trust guidance and LLM security** publications. The *Using Zero Trust to Secure Enterprise Information in LLM Environments* [9] translates Zero Trust architecture principles—least privilege access, continuous verification, and microsegmentation—into concrete recommendations for LLM and agentic AI

environments, including guidance on monitoring data flows, constraining agent permissions, and maintaining observability across AI system boundaries. Applying Zero Trust principles to AI systems provides the policy verification layer that allows anomalies in agent behavior to be detected against a defined baseline rather than through signature matching alone.

The governance structures and accountability models underlying the visibility deficit are directly addressed in **AI Organizational Responsibilities: Governance, Risk Management, Compliance and Cultural Aspects** [7]. This publication provides detailed guidance on establishing AI governance boards, defining ownership and accountability, developing AI security policies, and managing shadow AI through organizational and cultural means—including the combination of policy clarity, accessible sanctioned alternatives, and detection capability that the evidence identifies as most effective at reducing unsanctioned deployment. For organizations building AI security governance on top of existing cloud security programs, **CCM v4.1** provides the mature governance foundation on which AI-specific controls should be layered [11]. Its well-established risk management processes and control frameworks offer a practical starting point that can be extended to AI-specific requirements through AICM domain mapping.

Conclusions and Recommendations

The enterprise AI visibility crisis is real, measurable, and consequential. Two-thirds of CISOs cannot see how AI is deployed across their organizations. Three-quarters are relying on security tools that were not designed to monitor AI. One in five organizations has deployed AI agents with no monitoring or guardrails. The financial and security consequences of this exposure are documented and growing.

The path forward is governance-first. The evidence is consistent across multiple surveys and research streams: organizations with formal AI governance programs achieve demonstrably better security visibility than those without. Governance creates the structures—asset registries, deployment review processes, monitoring requirements, accountability assignments—that make visibility possible. Technology alone cannot close the gap; without the governance foundation, monitoring tools produce noise rather than insight.

Organizations addressing the visibility deficit should pursue the following priorities.

Establish a comprehensive AI asset inventory as an immediate priority, combining proactive discovery mechanisms with mandatory disclosure requirements for all new AI deployments. The inventory should capture not only formally approved systems but also shadow AI deployments identified through network analysis, procurement review, and departmental outreach.

Adopt the CSA AICM as the governance framework for AI security, using its 18 control domains to structure both the assessment of existing deployments and the requirements for new ones. Completing an AI-CAIQ self-assessment through the STAR for AI Registry provides a structured starting point and a mechanism for ongoing maturity measurement.

Invest in AI-specific monitoring capabilities, recognizing that infrastructure-layer monitoring cannot substitute for agent-layer observability. For agentic AI deployments, implement monitoring that captures the full decision chain—from prompt through reasoning to tool invocation and output—using MAESTRO's threat model to ensure coverage across the relevant attack surfaces.

Apply Zero Trust principles to AI system access, ensuring that AI agents operate under least-privilege constraints, that their permissions are regularly reviewed and rotated, and that their access to sensitive systems is mediated through explicit policy verification rather than inherited from broad trust relationships.

Build organizational expertise in AI security, recognizing that the skills deficit is a primary driver of the visibility gap. Security teams that do not understand AI-specific risks cannot design effective monitoring for them. Investment in training, certification, and dedicated AI security roles is a prerequisite for meaningful visibility capability.

Establish AI governance as a shared organizational function rather than a security team problem alone. The visibility deficit is a product of governance fragmentation; resolving it requires clear accountability structures that connect deployment authority to security oversight across all functions that deploy AI systems.

Organizations that close this visibility gap will not only reduce their security risk—they will be positioned to adopt AI at scale with the governance infrastructure required to sustain responsible deployment over time.

References

[1] Pentera, "AI Security and Exposure Benchmark Report 2026," February 18, 2026. Survey of 300 U.S. CISOs and senior security leaders. Available at: https://www.prnewswire.com/il/news-releases/ai-is-spreading-faster-than-companies-can-secure-it-ciso-survey-finds-302691361.html

[2] Help Net Security, "AI went from assistant to autonomous actor and security never caught up," March 3, 2026. Available at: https://www.helpnetsecurity.com/2026/03/03/enterprise-ai-agent-security-2026/

[3] Cloud Security Alliance and Google Cloud, "The State of AI Security and Governance," conducted Summer 2025, published 2025. Survey of 300 IT and security professionals. Available at: https://cloudsecurityalliance.org/artifacts/the-state-of-ai-security-and-governance

[4] Help Net Security, "Enterprises are racing to secure agentic AI deployments," February 23, 2026. Available at: https://www.helpnetsecurity.com/2026/02/23/ai-agent-security-risks-enterprise/

[5] ISACA, "From Shadow IT to Shadow AI: Navigating the New Frontier of Enterprise Risk," Volume 19, 2025. Available at: https://www.isaca.org/resources/news-and-trends/newsletters/atisaca/2025/volume-19/from-shadow-it-to-shadow-ai-navigating-the-new-frontier-of-enterprise-risk

[6] Cloud Security Alliance, "AI Gone Wild: Why Shadow AI Is Your IT Team's Worst Nightmare," March 4, 2025. Available at: https://cloudsecurityalliance.org/blog/2025/03/04/ai-gone-wild-why-shadow-ai-is-your-it-team-s-worst-nightmare

[7] Cloud Security Alliance, "AI Organizational Responsibilities: Governance, Risk Management, Compliance and Cultural Aspects," 2024. Authors: Nick Hamilton, Ken Huang, Michael Roza. AI Organizational Responsibility Working Group. Available at: https://cloudsecurityalliance.org

[8] Cloud Security Alliance, "AI Controls Matrix (AICM) v1.0 Introduction Guidance," 2025. AI Controls Framework Working Group. Lead Author: Marina Bregkou. Available at: https://cloudsecurityalliance.org

[9] Cloud Security Alliance, "Using Zero Trust to Secure Enterprise Information in LLM Environments," 2025. Available at: https://cloudsecurityalliance.org

[10] Cloud Security Alliance AI Safety Initiative, "MAESTRO: Multi-Agent Environment Security Threat Review and Operations," 2025. Available at: https://cloudsecurityalliance.org

[11] Cloud Security Alliance, "Cloud Controls Matrix (CCM) v4.1," 2021. Available at: https://cloudsecurityalliance.org/research/cloud-controls-matrix

[12] Cloud Security Alliance and Google Cloud, "The State of AI and Security Survey Report," 2024. Survey of 2,486 IT and security professionals. Authors: Hillary Baron et al. Available at: https://cloudsecurityalliance.org/artifacts/the-state-of-ai-and-security-survey-report

[13] Cloud Security Alliance, "AI Governance: A Maturity Multiplier," December 18, 2025. Available at: https://cloudsecurityalliance.org/blog/2025/12/18/ai-security-governance-your-maturity-multiplier

[14] IANS Research, "The CISO's Expanding AI Mandate: Leading Governance in 2026," February 6, 2026. Available at: https://www.iansresearch.com/resources/all-blogs/post/security-blog/2026/02/06/the-cisos-expanding-ai-mandate--leading-governance-in-2026

[15] Microsoft and LinkedIn, "2024 Work Trend Index Annual Report: AI at Work Is Here. Now Comes the Hard Part," May 2024. Available at: https://www.microsoft.com/en-us/worklab/work-trend-index

[16] Akto, "State of Agentic AI Security 2025," February 2026. Available at: https://www.akto.io/blog/state-of-agentic-ai-security-2025