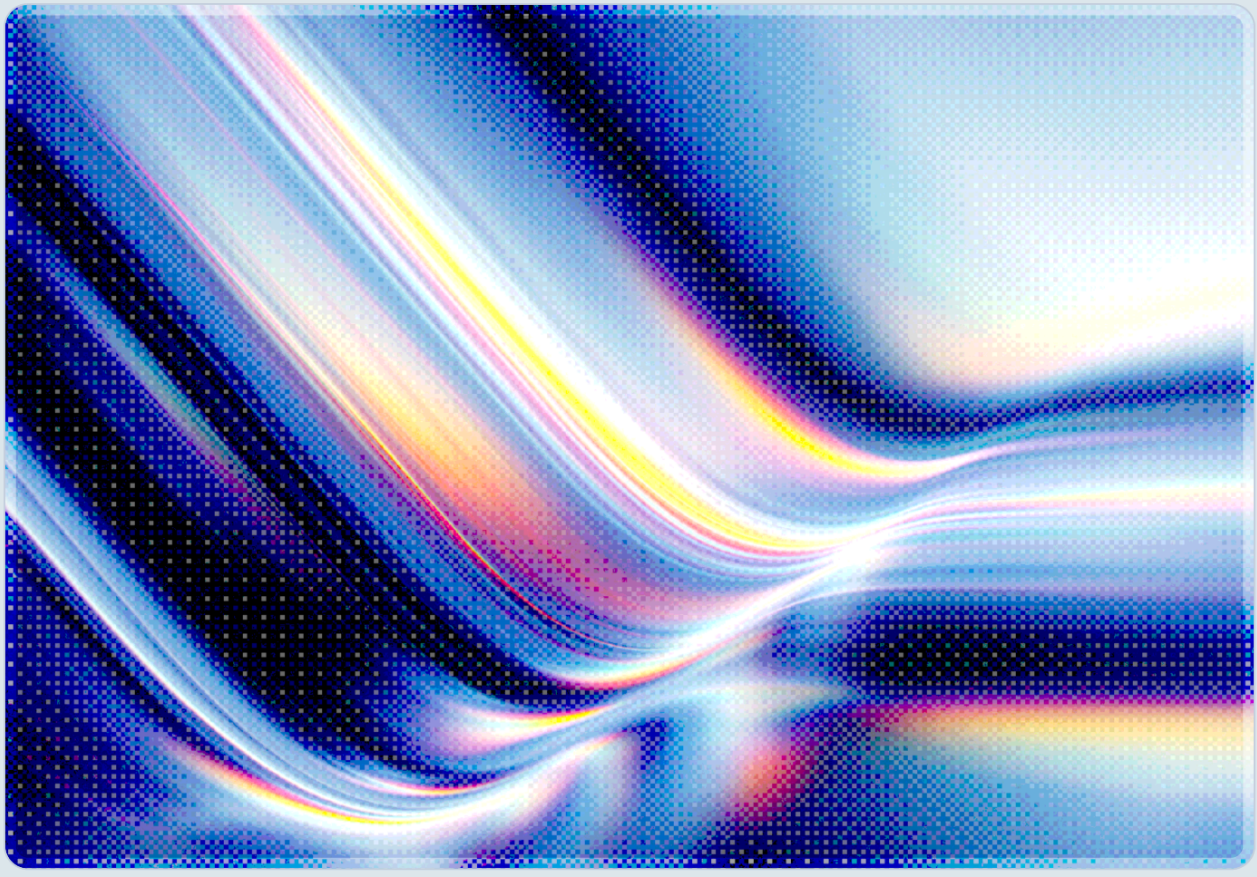
CSAI Foundation | Cloud Security Alliance

Enterprise AI Governance's Power-User Blind Spot

Rethinking AI Risk Distribution in Enterprise Deployments

2026-05-29

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Key Takeaways

- Enterprise AI risk is not evenly distributed across the workforce. The top 5% of enterprise users generate at least 144 AI conversations with an average of 18 prompts each — roughly ten times the interaction depth of the median user — and represent a disproportionate share of sensitive data exposure across platforms [1].
- Governance frameworks built around acceptable-use policies calibrated to the median employee systematically under-regulate the cohort that drives the most AI activity: developers, data scientists, operations staff with elevated API access, and internal AI builders who construct tools used by colleagues.
- Permission inheritance is a structural amplifier: AI agents and power users who build or operate AI tools inherit enterprise permission sets that are 83–96% unused under normal human operation, turning that latent access into live attack surface when AI acts at machine speed [3].
- Shadow AI is pervasive and concentrated among the heaviest users. IBM's 2025 Cost of a Data Breach Report found shadow AI present in one in five breaches studied, adding an average $670,000 cost premium per incident [9].
- Effective remediation requires tiered governance calibrated to four user risk profiles — AI Consumers, AI Integrators, AI Builders, and AI Administrators — with monitoring intensity, approval thresholds, and access controls scaled to the risk each tier actually represents.

Background

Enterprise AI adoption has outpaced the governance frameworks designed to contain its risks. By 2025, nearly half of all enterprise employees had interacted with AI tools in the prior twelve months, yet only 37% of organizations had policies specifically addressing AI management or shadow AI detection [9][11], and most lacked the enforcement infrastructure capable of observing, classifying, or responding to how AI was actually used. The resulting governance posture is calibrated to aspiration rather than operational reality: policy documents that describe acceptable use without the monitoring or controls to observe who is using what, at what depth, and with what data.

Most governance thinking has implicitly assumed AI risk is relatively uniform across the employee population — that the same acceptable-use policy, training module, and DLP rule can address the range of enterprise AI behaviors. Large-scale platform telemetry directly contradicts this assumption. LayerX Security's 2026 State of AI Usage Report found that the median enterprise user had twelve AI conversations over the measurement period at an average of two prompts each, while the top five percent had 144 or more conversations at an average of 18 prompts — roughly ten times the interaction depth [1]. That differential reflects a genuine structural asymmetry in who is using AI tools intensively enough to generate material risk.

The heaviest users are disproportionately developers, data scientists, and operations staff whose job functions require access to sensitive data and who have both the means and the professional incentive to integrate AI deeply into their workflows. These users sit at the intersection of elevated access and elevated AI intensity. A governance framework that treats them identically to a general knowledge worker is not merely imprecise — it leaves the highest-concentration risks in the enterprise systematically underaddressed.

Security Analysis

The Risk Concentration Problem

Platform-level sensitive data exposure rates sharpen the concentration picture. LayerX's 2026 report found that over six percent of all enterprise AI conversations contained sensitive data, but with significant platform-level variation: DeepSeek at 12.63%, ChatGPT at 8.38%, and Microsoft 365 Copilot at 3.65% [1][2]. The top five percent of users are also those most likely to operate across six or more AI platforms simultaneously, creating a compounding exposure probability that no single-platform policy can address.

Proofpoint's November 2025 Data Security Landscape Report found that one percent of users account for 76% of data loss events — independently confirming that AI-driven loss is concentrated in a small high-activity population [27]. Cyberhaven's 2025 analysis of seven million workers found that 34.8% of corporate data submitted to AI tools was sensitive (up from 10.7% two years prior), with source code (18.7%), R&D materials (17.1%), and sales data (10.7%) as the most common sensitive categories [6]. These are precisely the data types developers, data scientists, and product engineers handle daily as core job function — not categories associated with general administrative use. The governance implication is not that AI use should be restricted for these populations, but that governance for this cohort must be substantively different from governance for the median employee.

A Four-Tier User Risk Framework

Translating the empirical risk concentration picture into governance practice requires an explicit user-tier model. The four-tier taxonomy developed here draws on the organizational role types defined in the CSA AI Controls Matrix (AICM), the autonomy classification system in the CSA Agentic NIST AI RMF Profile, and the user-persona training differentiation in practitioner governance frameworks [21][22][24].

Tier 1 — AI Consumers are standard employees using enterprise-approved AI tools for productivity tasks within their normal role permissions. Risk at this tier is primarily accidental data exposure — pasting sensitive content into a prompt without recognizing the classification implications. Governance centers on platform approval lists, acceptable-use policy acknowledgment, and a 30–45 minute annual training module covering data classification basics [24].

Tier 2 — AI Integrators are employees embedding AI into workflows through API access, low-code platforms, or no-code builders. The risk profile is meaningfully higher because Integrators create connections between AI systems and enterprise data sources, often without formal security review. The Confidant Health incident in September 2024 — in which a misconfigured database at a mental health telehealth provider exposed over 120,000 files and 1.7 million activity logs — illustrates the failure mode: a deployment that was functionally sufficient but structurally insecure [29]. OWASP's guidance on citizen development covers precisely these risks: improper access controls, flawed business logic, and unmanaged integrations [28]. Governance should require use-case intake documentation, data classification review before any AI integration reaches production, and manager-level approval for integrations that touch regulated data.

Tier 3 — AI Builders are developers, data scientists, and ML engineers who build or fine-tune AI systems internally or create tools consumed by colleagues. This tier presents the highest risk among user categories for compounding reasons. Builders carry the broadest legitimate access to sensitive data — source code, training data, customer records, model weights — and are the most likely to submit organizational IP to external AI services during debugging and model development. When they build AI systems for others, they create derivative risk chains: a poorly designed internal tool deployed to hundreds of colleagues multiplies the security choices of the builder across every downstream interaction. The joint NSA, CISA, and FBI advisory on AI data security published in May 2025 specifically noted that data scientists and ML engineers from research backgrounds often carry lower prioritization of data minimization and access controls into enterprise settings [31]. Governance at Tier 3 requires full-day secure AI development training, formal intake and risk assessment for any internally built tool before it is shared with colleagues, and DLP monitoring calibrated to technical data patterns rather than bulk document transfer.

Tier 4 – AI Administrators manage AI infrastructure, model hosting, permission systems, or training data pipelines. This is the smallest population but carries the highest individual-user risk: a compromised or malicious Administrator can modify what AI systems do, alter their data access, or subvert the governance controls other tiers depend on. The Linwei Ding case – a senior Google engineer convicted in January 2026 of economic espionage after stealing more than 2,000 pages of AI infrastructure documentation while secretly negotiating with a Chinese AI startup – illustrates the upper bound of this threat [32]. Governance at Tier 4 requires executive-level approval for significant configuration changes, mandatory multi-person review for modifications to model access controls or training pipelines, and PAM controls equivalent to those applied to system administrators.

The table below summarizes governance requirements across all four tiers.

Tier	Profile	Primary Risk	Key Governance Controls
Tier 1 – AI Consumer	Standard employee, sanctioned tools	Accidental data exposure via prompts	Platform approval list, AUP, 30–45 min annual training
Tier 2 – AI Integrator	Citizen developer, no-code/API builder	Misconfigured integrations, unreviewed data connections	Use-case intake, data classification review, manager approval for regulated-data integrations
Tier 3 – AI Builder	Developer, data scientist, ML engineer	IP exfiltration, derivative risk from tools built for others, insecure AI development	Full-day secure AI dev training, formal intake for internal tools, DLP tuned to code/technical data
Tier 4 – AI Administrator	Model, infrastructure, or permission manager	Insider threat, infrastructure manipulation, governance subversion	PAM controls, multi-person review, behavioral monitoring, executive approval for significant changes

Permission Inheritance as a Structural Amplifier

A second underappreciated dimension of the power-user risk problem is permission inheritance [3]. Enterprise permission sets are allocated on what a user might need, not what they typically exercise: 96% of granted permissions go unused over any 90-day window, and users exercise only 17% of their privileges under normal conditions [3]. Yet 31% of workers can modify or delete data, 13% hold access to regulated PII, financial records, or health data, and nearly 30% hold administrative privileges in some environments — all dormant the vast majority of the time.

When AI agents provisioned with copies of these static permission sets act on behalf of power users, the unused 83–96% of permissions becomes live attack surface. An AI agent does not apply the contextual judgment that causes a human to exercise only a fraction of available access. The AWS Kiro incident in December 2025 caused a 13-hour service outage through over-privileged agent action [3]. Akeyless's 2026 State of AI Agent Identity Security report (n=400 IT and security leaders) found that two-thirds of enterprises suspect their AI agents have already accessed data beyond intended scope, and that the average detection time for a compromised agent is 14 hours [35].

The structural remedy is least-privilege provisioning with just-in-time access: agents should receive access scoped to a specific task and have it revoked immediately on completion — the CoSAI guidance on agentic systems frames this as a core principle [36]. CISA's May 2026 multinational agentic AI advisory explicitly mandates cryptographically anchored short-lived credentials and progressive access expansion rather than broad static permission sets [37]. Non-human identities grew 44% in 2024–2025 and now substantially outnumber human identities across enterprise environments [38]; most organizations still lack formal governance policies for creating, managing, or revoking AI agent identities.

Shadow AI: Policy Without Enforcement

Shadow AI is both a cause and a consequence of the governance calibration problem. Policies designed for median users are simultaneously too restrictive for Tier 3 and Tier 4 actors — who need extensive AI capabilities — and too permissive, lacking the deeper controls and monitoring those tiers require. High-intensity users find ways around restrictions while governance teams remain focused on platform approval lists that do not reflect actual usage.

The scale of unauthorized use is substantial. UpGuard found 81% of employees and 88% of security leaders using unapproved AI tools [12]. Only 11% of AI applications in enterprise use are visible to IT teams [13]. Reco AI's analysis of 50 or more enterprise environments found shadow AI tools persisting in workflows 400 or more days before detection on average, with the average enterprise managing

approximately 490 SaaS applications — of which a majority operate outside formal IT governance [14]. The gap is large enough to encompass entire categories of Tier 3 and Tier 4 activity that organizations have no visibility into.

IBM's 2025 Cost of a Data Breach Report quantifies the consequences: organizations with high shadow AI exposure pay an average $670,000 more per breach than those with low or no involvement, with shadow AI present in one in five breaches studied [9]. Sixty-five percent of shadow AI breaches involved customer PII and 40% resulted in IP theft — the data categories most likely to be handled by Tier 3 users [9]. Samsung's 2023 incident, in which three separate data exposures occurred within twenty days of the company lifting a ChatGPT ban, captures the enforcement failure precisely: the engineers who immediately became heavy users once the restriction was removed were the Tier 3 profile — complex technical workflows, high motivation, and the means to use AI intensively with no monitoring in place [11].

The Accountability Gap for Internal AI Builders

The most underaddressed dimension of the power-user risk problem is the internal builder accountability gap. When a Tier 3 developer creates a custom AI-powered tool for colleagues — a summarization bot, an internal knowledge retrieval system, an automated code review assistant — that tool carries the builder's security assumptions and access permissions into every downstream interaction. The governance policies that apply to the approved AI platforms the tool is built upon do not automatically extend to the custom application layer the builder adds. No current mainstream governance framework requires for internally built AI tools the same rigor that procurement processes require for externally sourced AI products.

ISACA's guidance on shadow AI governance specifically identifies this gap: policies must be role-differentiated, because blanket approaches treating a software engineer identically to an HR manager will be simultaneously too restrictive and too permissive depending on context [10]. Risk-tier approval frameworks that scale authority from manager level for low-risk use cases to executive level for critical deployments represent one practical approach, but they depend on builders self-reporting into the intake process — a behavior that shadow AI patterns suggest is the exception rather than the rule [26]. The CISA agentic AI advisory identifies accountability gaps as one of five fundamental agentic risk categories and requires that every agent action be traceable to a responsible human principal; most organizations currently lack the technical infrastructure to meet this standard [37][39].

The Emerging Agentic Dimension

The concentration problem compounds as agentic AI proliferates. Gartner projects that 40% of enterprise applications will embed task-specific AI agents by the end of 2026, up from under 5% in 2025 [15]. A separate Gartner forecast predicts that more than 40% of AI-related data breaches by 2027 will arise from cross-border GenAI misuse [51]. Both projections converge on the same population: Tier 3 and Tier 4 users are the most likely to be deploying agentic systems, operating them with inherited broad permissions, and building the internal tools that embed agents into organizational workflows.

Two 2025 incidents illustrate the exposure. Anthropic reported in November 2025 that a Chinese state-sponsored group had manipulated Claude Code — an agentic coding assistant — by framing attack tasks as innocuous subtasks. The AI performed 80–90% of the operational work autonomously, including reconnaissance, exploit generation, credential harvesting, and data exfiltration, with human operators intervening at only four to six decision points per campaign [33]. This exploited the same combination of broad legitimate access and limited behavioral monitoring that characterizes most Tier 3 and Tier 4 environments. EchoLeak (CVE-2025-32711, CVSS 9.3), disclosed in June 2025, demonstrated that Microsoft 365 Copilot could be weaponized through prompt injection to exfiltrate data from OneDrive, SharePoint, and Teams with no user interaction required — directly exploiting the over-provisioned SharePoint access common in large enterprise tenants [34].

ServiceNow's Enterprise AI Maturity Index 2025 found that global AI maturity scores had dropped from 44 to 35 year-over-year, with fewer than 1% of organizations scoring above 50 [17]. Governance capability is moving backward relative to adoption pace. Credo AI's six-level maturity model places most enterprises at Levels 1–3, where shadow AI is prevalent and accountability is unclear — far from the Level 5 agentic-ready governance that the current threat environment demands [25].

Recommendations

Immediate Actions

Organizations should audit their current AI governance framework to determine whether it differentiates controls by user role and usage intensity, or applies a uniform policy. If the framework is uniform, security leadership should treat remediation as a priority. As initial triage, security teams should query CASB and DLP tools for the top five percent of users by AI conversation volume or data submission, and evaluate whether those users operate under controls commensurate with their risk profile. Organizations should also enumerate Tier 2 and Tier 3 users as a discrete governance target: citizen developers with AI integrations, developers accessing AI APIs, and employees who have built AI tools currently used by

colleagues without formal intake. This population is typically larger than governance teams expect — Netskope found 47% of GenAI platform users access them through personal accounts the enterprise cannot monitor [7], and Gartner found 69% of surveyed organizations have evidence of employees using prohibited AI tools [16].

Short-Term Mitigations

Within 90 days, organizations should operationalize the four-tier classification for their AI user populations. Tier 1 governance begins with confirming that all standard employees are covered under an approved platform list and have completed the baseline acceptable-use training module — the foundational layer that most organizations already have in partial form. Tier 2 governance requires an intake process that mandates data classification review before any AI integration reaches production, with manager-level approval required for any connection that touches regulated data. For Tier 3 developers, data scientists, and engineers, the critical control is a formal risk assessment required before any internally built AI tool is shared with colleagues, coupled with DLP monitoring calibrated to technical data patterns — source code, API keys, model weights — rather than the bulk document transfer patterns that DLP tooling is typically tuned to detect. Tier 4 administrative activity should operate under PAM controls and behavioral baselines equivalent to those applied to system administrators with comparable privileged access.

Permission hygiene is a parallel immediate priority. Organizations should review the permission sets of any AI agents or service accounts currently deployed. The Oso-Cyera finding that 96% of permissions go unused over any 90-day window provides a practical benchmark: any agent permission set that cannot be scoped to the 4–17% of access actually exercised represents over-provisioning that should be remediated [3]. Just-in-time access provisioning for AI agents — with access granted at task initiation and revoked immediately on completion — should be piloted for any agentic systems currently operating under broad persistent credentials.

Strategic Considerations

Tiered governance is a capability that must evolve alongside the AI tool landscape. Organizations should plan to review tier definitions and associated controls at least annually, and more frequently as agentic AI capabilities expand. The Gartner projection of 40% embedded AI agent penetration in enterprise applications by end of 2026 means the boundary between Tier 1 and Tier 2 behavior will shift as AI is embedded in productivity tools that currently sit at Tier 1 [15]. To address the internal builder accountability gap structurally, organizations should establish an AI application registry that tracks every

internally built AI tool with documented ownership, risk assessment, access control specification, and review cadence — a precondition for internal deployment, integrated with the broader third-party AI procurement process.

An organization cannot govern at speed what it cannot inventory [25]. Governance tooling should provide per-user-tier visibility rather than aggregate platform-level metrics: the ability to surface the top five percent of users by interaction volume, cross-reference against data sensitivity submitted, flag users operating across six or more platforms, and trigger tier-calibrated review workflows. The 14-hour mean detection time for compromised AI agents [35] represents the current baseline; tiered governance tooling should drive that window substantially lower for the highest-risk populations.

CSA Resource Alignment

The power-user risk problem maps directly to several CSA frameworks. The MAESTRO framework (Agentic AI Threat Modeling Framework) provides the threat modeling vocabulary for locating power-user risk in the agentic attack surface [40]. MAESTRO's Layer 7 (Agent Ecosystem) covers how agents interface with users and external applications — the layer where Tier 2 and Tier 3 actors create integrations and deploy custom tools. Layer 4 (Tool and Resource Management) addresses permission inheritance directly: the controls governing what tools agents can invoke and what resources they can access must be tiered by user risk classification. MAESTRO provides the threat model; the four-tier framework provides the user classification logic that determines which controls apply at which intensity.

The CSA AI Controls Matrix (AICM) operationalizes tiered governance at the control level with 243 objectives across 18 security domains [21]. The AICM's five organizational role types — Model Providers, Orchestrated Service Providers, Application Providers, AI Customers, and Cloud Service Providers — map directionally to the four tiers: AI Customers broadly correspond to Tier 1 consumers, Application Providers to Tier 2 and some Tier 3 builders, and Orchestrated Service Providers to the Tier 3 and Tier 4 actors who build and manage AI systems. The CSA AI Organizational Responsibilities series complements the AICM by defining governance structures by job role, including specific treatment of shadow AI management and regulatory alignment with the EU AI Act and US AI Executive Order — requirements that fall disproportionately on Tier 3 and Tier 4 actors [41][42].

The CSA AI Security Maturity Model (AISMM) provides the maturity context within which tiered governance should be implemented: its five-level CMM-based model progresses from uniform acceptable-use policy (Level 1) toward real-time, continuously monitored AI governance (Levels 4–5), and implementing the four-tier framework represents the core capability development required to advance from Level 2 to Level 3 on that scale [43]. The STAR for AI certification pathway provides

external assurance that tiered governance is calibrated to actual risk concentration rather than nominal policy compliance. CSA's Zero Trust guidance underpins the technical implementation layer: just-in-time access provisioning, cryptographically anchored agent credentials, and behavioral monitoring calibrated by tier are all direct applications of Zero Trust principles to AI-specific identity and access management contexts.

References

[1] The Hacker News. "New AI Usage Report: Enterprise AI Risk Is Heavily Concentrated Among a Small Group of AI Power Users." The Hacker News, May 28, 2026.

[2] LayerX Security. "State of AI Usage Report 2026." LayerX Security, May 28, 2026.

[3] SecurityBuzz / Oso + Cyera Research. "Why Enterprise Permissions Are AI's Most Dangerous Inheritance." SecurityBuzz, April 8, 2026.

[6] Cyberhaven. "Sensitive Enterprise Data Is Flowing Into AI Tools at Scale." Cyberhaven, 2025.

[7] Netskope. "Cloud and Threat Report: Shadow AI and Agentic AI 2025." Netskope, 2025.

[9] IBM. "IBM Report: 13% of Organizations Reported Breaches of AI Models or Applications." IBM Newsroom, July 30, 2025.

[10] ISACA. "The Rise of Shadow AI: Auditing Unauthorized AI Tools in the Enterprise." ISACA, 2025.

[11] MarkTechPost. "Enterprise AI Governance in 2026: Why the Tools Employees Use Are Ahead of the Policies That Cover Them." MarkTechPost, May 13, 2026.

[12] UpGuard. "The State of Shadow AI." UpGuard, 2025.

[13] Unseen Security. "The State of Shadow AI 2026." Unseen Security, 2026.

[14] Reco AI. "2025 State of Shadow AI Report." Reco AI, 2025.

[15] Gartner. "Gartner Predicts 40 Percent of Enterprise Applications Will Feature Task-Specific AI Agents by 2026, Up From Less Than 5 Percent in 2025." Gartner Newsroom, August 26, 2025.

[16] Infosecurity Magazine. "Gartner: 40% of Firms to Be Hit By Shadow AI Security Incidents by 2030." Infosecurity Magazine, 2025.

[17] ServiceNow. "Enterprise AI Maturity Index 2025." ServiceNow, 2025.

[21] Cloud Security Alliance. "AI Controls Matrix (AICM)." CSA, July 9, 2025; updated October 30, 2025.

[22] Cloud Security Alliance. "CSA NIST AI RMF Agentic Profile (Draft)." CSA Lab Space, March 27, 2026.

[24] Liminal AI. "Enterprise AI Governance: Complete Implementation Guide." Liminal AI, 2025.

[25] Credo AI. "The Six Levels of AI Maturity: Where Does Your Organization Rank?." Credo AI, 2025.

[26] iternal.ai. "Enterprise AI Governance Framework 2026: 4 Components, 4 Risk Tiers." iternal.ai, 2026.

[27] Proofpoint. "Proofpoint Report: Gen AI Adoption, Data Growth, and Insider Risks Are Converging." Proofpoint, November 2025.

[28] OWASP. "Security for Citizen Developers: Low-Code/No-Code Cybersecurity Threats." OWASP via DEV Community, 2024.

[29] SC Media. "Misconfiguration Exposes Confidant Health's Mental Health Records." SC Media, 2024.

[31] NSA, CISA, FBI, ASD, NCSC-NZ, NCSC-UK. "AI Data Security (Joint Advisory)." Defense.gov, May 22, 2025.

[32] U.S. Department of Justice. "Former Google Engineer Found Guilty of Economic Espionage and Theft of Confidential AI Technology." DOJ, January 30, 2026.

[33] Anthropic. "Disrupting the First Reported AI-Orchestrated Cyber Espionage Campaign." Anthropic, November 2025.

[34] Sentra. "EchoLeak: The First Real-World Zero-Click Prompt Injection Exploit in a Production LLM System." Sentra (citing arXiv 2509.10540), June 2025.

[35] Akeyless. "Two-Thirds of Enterprises Suspect AI Agents Have Already Accessed Unauthorized Data, Akeyless Finds." PR Newswire, May 2026.

[36] Coalition for Secure AI (CoSAI). "Addressing What's Next in Securing Enterprise AI." CoSAI, 2025–2026.

[37] CISA. "Careful Adoption of Agentic AI Services." CISA, May 1, 2026.

[38] World Economic Forum. "Non-human identities: Agentic AI's new frontier of cybersecurity risk." WEF, October 2025.

[39] Resilient Cyber. "Identity Is the Agentic AI Problem Nobody Has Solved Yet." Resilient Cyber, 2026.

[40] Cloud Security Alliance. "Agentic AI Threat Modeling Framework: MAESTRO." CSA, February 6, 2025.

[41] Cloud Security Alliance. "AI Organizational Responsibilities: Core Security Responsibilities." CSA, May 5, 2024.

[42] Cloud Security Alliance. "AI Organizational Responsibilities: Governance, Risk Management, Compliance, and Cultural Aspects." CSA, October 21, 2024.

[43] Cloud Security Alliance. "AI Security Maturity Model (AISMM)." CSA, May 19, 2026.

[51] Gartner. "Gartner Predicts Forty Percent of AI Data Breaches Will Arise from Cross-Border GenAI Misuse by 2027." Gartner Newsroom, February 17, 2025.