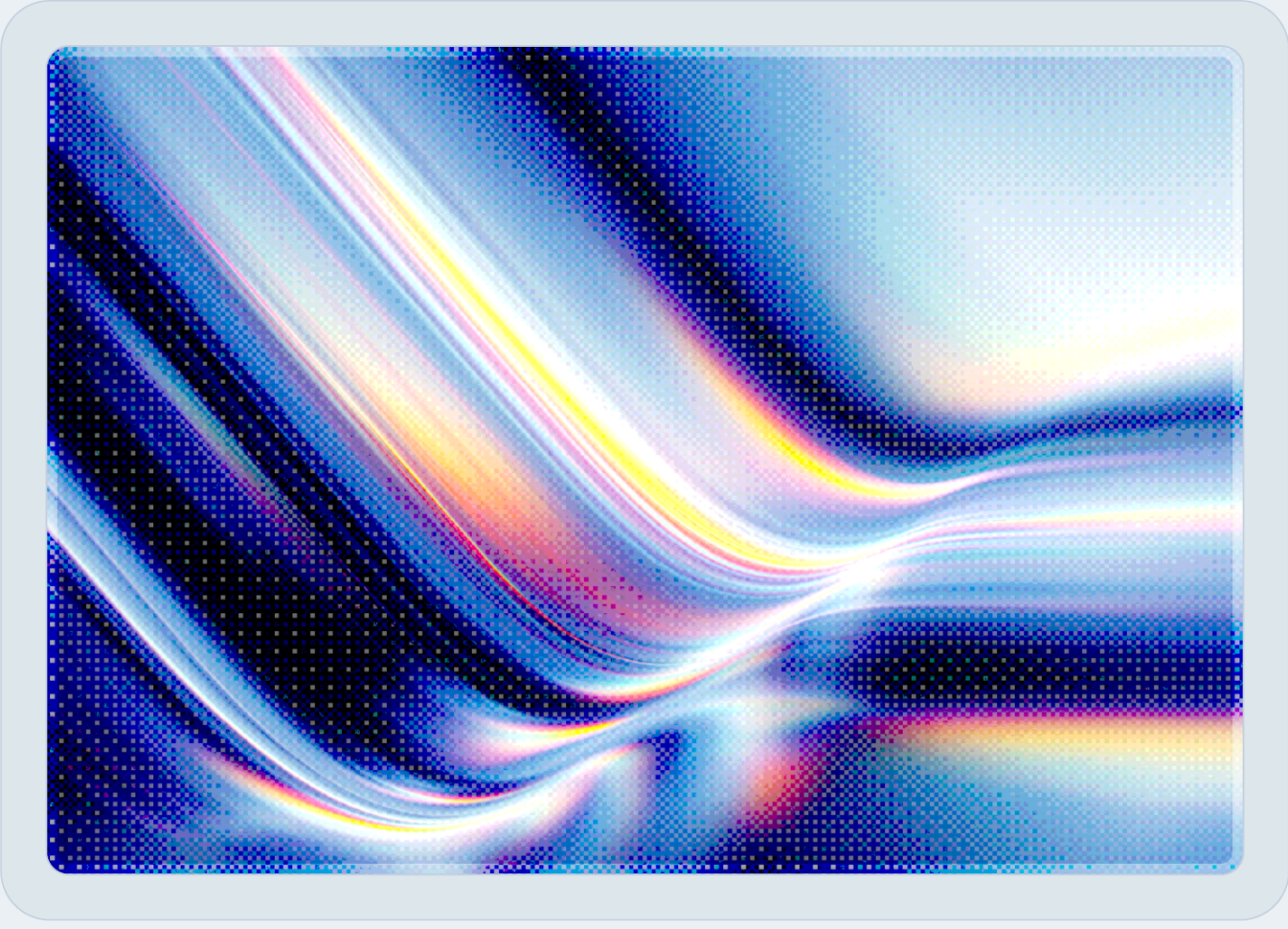
CSAI Foundation | Cloud Security Alliance

Undeclared AI Usage: Cyber Insurance's Emerging Concentration Risk

2026-09-21

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Key Takeaways

- Employee use of generative AI on corporate devices has tripled year over year to roughly 45%, with 67% of that usage occurring through personal accounts outside IT's visibility — a gap that suggests insurers have limited visibility into which AI tools and models a given policyholder actually depends on [1][4].
- Industry estimates put enterprise AI usage at somewhere between 60% and 80% concentrated on a small handful of frontier foundation models, meaning a single provider's outage, safety regression, or regulatory restriction can generate correlated losses across many policyholders at once — an accumulation dynamic comparable to what makes catastrophe risk hard to price [1][10].
- Cyber insurers appear to be moving toward converting "silent AI" — implicit, undeclared AI exposure sitting inside ordinary cyber, technology E&O, and general liability policies — into explicit exclusions and harder-to-resolve attribution disputes, based on current market-hardening trends and the attribution challenges AI-generated malware poses to war-exclusion clauses, while affirmative AI coverage remains small and unevenly available [3][11][13].
- Gartner projects that more than 40% of organizations will experience a security or compliance incident tied to unauthorized AI tools by 2030, and 42% of surveyed property and casualty insurers admit they have not yet measured the outcomes of their own AI deployments — underscoring that both sides of the underwriting relationship are working with incomplete visibility [5][6].
- Treating a foundation model as a swappable component — decoupled from an organization's data, retrieval, permissions, and validation logic — is a concrete technical mitigation available today, and it directly supports the kind of AI asset inventory that underwriters increasingly expect at renewal [2][9].

Background

Cyber insurers built their underwriting models on the assumption that a policyholder's technology stack is knowable: a defined set of vendors, cloud providers, and software dependencies that can be inventoried, questioned, and priced. Generative AI has undermined that assumption. According to

Verizon's 2026 Data Breach Investigations Report, regular employee use of AI tools on corporate devices has roughly tripled over the past year to approximately 45%, and two-thirds of that usage flows through personal, non-corporate accounts that never touch an organization's approved technology list [4]. KYND, a cyber risk analytics firm, framed the resulting problem for insurers during a September 1, 2026 industry webinar building on its white paper "The Wild West of AI Risk": businesses are wiring AI into hiring, customer service, claims handling, and dozens of other workflows faster than their insurers can ask about it [1][8].

The visibility gap compounds a second, related structural problem: the AI tools employees adopt, declared or not, are themselves drawn from a narrow pool of underlying providers. Industry estimates put the share of enterprise AI usage running on a small cluster of frontier models at somewhere between 60% and 80% [1]. CSA's own research has separately identified this same dynamic — foundation-model concentration converted into an insurability problem — terming it "silent-AI" exposure and recommending aggregation limits by model provider as a mitigation [13]. A supply-chain analysis published by Logistics Viewpoints in mid-September 2026 makes a related point from the enterprise-architecture side: manufacturers building transportation, procurement, planning, and warehousing workflows around a single foundation model's behavior create a dependency that is difficult to unwind once operational processes, prompts, and integrations are tuned to that model's quirks, even though switching providers is theoretically straightforward [2]. The article notes that Anthropic CEO Dario Amodei's call to "pace the frontier" of AI development with safety measures drew a pointed rebuttal from FTC Chairman Andrew Ferguson, who warned that compliance-heavy safety regimes can function as barriers that entrench incumbent model providers rather than genuinely reducing risk — a dynamic that, if it plays out, would deepen rather than ease the concentration problem insurers are now confronting [2].

Munich Re's "Cyber Insurance: Risks and Trends 2026" situates both dynamics inside a broader hardening of the supply-chain threat landscape: more than two-thirds of large organizations reported at least one third-party cybersecurity incident in the preceding twelve months, and nearly nine in ten C-level executives say their organization remains inadequately protected against cyberattacks [3]. Munich Re also reports that claims are shifting composition, with first-party losses now representing 62% of claims against 38% third-party; malicious incidents still outnumber non-malicious ones by roughly 3-to-1 in volume, but Munich Re describes non-malicious incidents — a category into which this note's authors would place most ungoverned AI usage, though Munich Re's report does not itself draw that connection — as growing in significance faster than malicious ones [3]. None of this is occurring in a vacuum for insurers specifically: Capgemini's World Property and Casualty Insurance Report 2026 found that 42% of P&C insurers have not measured the outcomes of their own AI initiatives, while only about 10% of the industry has successfully scaled AI into core operations [6]. The institutions being asked to price undeclared AI risk in their policyholders' operations are, in a meaningful number of cases, still working out how to govern AI in their own.

Security Analysis

The mechanics of undeclared AI exposure differ from most cyber risk categories because the loss driver is rarely a discrete technical failure. IBM's 2026 Cost of a Data Breach Report found that roughly one in four malicious breaches now involve an AI-enabled element — chiefly deepfake-assisted impersonation and AI-generated malware — and that these breaches cost organizations an average of $6 million, about $1 million above the $4.99 million global average across all breach types [7]. That severity premium is significant, but it likely understates the underwriting problem: most undeclared-AI losses probably will not present as a clean "AI breach" at all, since AI more often functions as an instrumentality within a broader loss than as an identifiable, standalone cause. A hallucinated compliance answer that triggers a regulatory fine, an autonomous agent that authorizes a fraudulent payment, or a vendor's AI feature that silently changes how customer data is processed are all losses in which AI functions as an instrumentality rather than an independently identifiable attack vector, which makes them harder for underwriters to price using conventional breach-based loss models.

Concentration turns this measurement problem into an accumulation problem. CSA's research note on OpenAI's July 25, 2026 outage — which simultaneously disabled ChatGPT, its API, and Codex, the latest in a pattern of roughly eighteen outages a month since January 2025 — used the incident as a live illustration of exactly this dynamic: enterprises dependent on a single foundation model provider have no independent path to the underlying capability when it fails, so a single upstream disruption produces correlated, simultaneous losses across every dependent organization rather than the independent, diversifiable losses that traditional actuarial pricing assumes [10]. With that provider alone serving more than 900 million weekly active users, the same logic applies to a discovered vulnerability, a safety regression, or a regulatory restriction affecting a widely deployed model — any of which could generate a correlated loss event across an insurer's book that looks nothing like a typical single-policyholder claim [10].

Insurers are also grappling with how AI complicates the attribution and causation questions that determine whether a loss is covered at all. A CSA research note on cyber insurance war exclusions found that AI-generated malware families capable of mimicking other threat actors' signatures and techniques are undermining the forensic attribution methods that war-exclusion clauses such as the London Market Association's LMA5567A depend on, making coverage disputes more likely and more protracted precisely when a loss involves AI-assisted tooling [11]. Taken together, these dynamics point toward a market narrowing coverage through both new exclusion language and harder-to-resolve attribution disputes, rather than developing the affirmative products and clearer definitions that would let policyholders know in advance what undeclared AI usage actually costs them in a claim.

Recommendations

Immediate Actions

Security and risk teams should treat the current renewal cycle as a practical deadline for building a defensible AI inventory: a documented list of which foundation models and AI-enabled vendor features the organization relies on, including shadow AI usage discovered through browser telemetry, SaaS-access reviews, and expense-report audits of AI subscriptions rather than self-reported IT lists alone. CSA's whitepaper "The Invisible Enterprise: Shadow AI and the Ungoverned Frontier" found that 91% of AI tools in enterprise environments operate outside IT control and that only 42% of organizations fully understand their own AI inventory — a governance gap that maps directly onto the visibility gap insurers are now pricing around [9]. Legal and risk teams should also request each carrier's written position on how AI-instrumented losses are treated under current cyber, CGL, and technology E&O wording — expressly covered, expressly excluded, or silent — rather than assuming last year's policy language still applies [3][11].

Short-Term Mitigations

Organizations should map their AI dependency concentration the same way they would map a single-supplier or single-region concentration risk: identifying which business-critical workflows depend on one foundation model provider, and where a provider outage, safety regression, or access restriction would create correlated failure across multiple functions simultaneously. Where feasible, architecture teams should decouple organizational data, retrieval logic, permissions, and validation from any single model's API surface, so that a provider substitution is an operational adjustment rather than a system rebuild — the same discipline the Logistics Viewpoints analysis recommends for AI-dependent supply chains [2]. CSA's CBRA-style approach of scoring AI systems on criticality, autonomy, permission scope, and potential impact offers a practical starting framework for prioritizing which shadow AI usage to remediate first [9].

Strategic Considerations

Boards and risk committees should treat undeclared AI usage as a governance failure with direct balance-sheet consequences, not merely an IT policy question, given that both the insurability of AI risk and the price of risk transfer now depend on an organization's ability to demonstrate what AI it actually runs. Enterprises with material AI concentration exposure should evaluate whether contractual continuity and audit provisions with model providers, or diversification across multiple foundation models for critical

workflows, offer more reliable protection than a traditional insurance market still calibrating how to price a risk it cannot yet fully see [3][10]. Over the longer term, organizations should watch for AI dependency disclosure emerging as a standard underwriting requirement, plausibly following a path similar to how vendor- and cloud-concentration disclosures have become more common in other commercial lines.

CSA Resource Alignment

This note builds directly on CSA's whitepaper "The Invisible Enterprise: Shadow AI and the Ungoverned Frontier," which is CSA's most specific prior work on the visibility failure underlying undeclared AI's insurance blind spot [9]. That paper's central finding — that 91% of AI tools in enterprise environments operate outside IT control, and that fewer than half of organizations fully understand their own AI inventory — is the root cause this note connects to the insurance market's growing unwillingness to write silent AI coverage. Its Capabilities-Based Risk Assessment framework, which scores AI systems on criticality, autonomy, permission scope, and potential impact, gives organizations a concrete method for building the AI inventories that underwriters are beginning to ask for at renewal [9].

This note also extends CSA's research note "Foundation Model Concentration: The Uninsurable AI Risk," published July 6, 2026, which first argued that foundation-model concentration is a structurally novel insurability problem and introduced the "silent-AI" exposure framing this note builds on, including its recommendation that insurers apply aggregation limits by model provider [13]. Where that note focused on the insurability question itself, this note connects it to the separate but related visibility problem of undeclared and shadow AI usage, and to the specific market behaviors — exclusion language and attribution disputes — now emerging in response to both.

CSA's research note "The ChatGPT Outage Pattern: Concentration Risk in Practice" supplies the concentration-risk half of this note's argument, documenting how dependency on a single foundation model provider converts an ordinary vendor outage into a correlated loss event across every dependent enterprise simultaneously — precisely the accumulation dynamic that makes undeclared AI usage hard for insurers to price as an independent risk [10]. CSA's research note "Attributing AI Attacks: When Cyber Coverage Becomes Conditional" extends the analysis to the claims side, showing how AI-generated malware's ability to mimic other threat actors undermines the attribution evidence that war exclusions and coverage determinations depend on [11]. Together, these four papers describe undeclared AI usage and foundation model concentration as two faces of the same underwriting problem: insurers cannot price what they cannot see, and what they can see is harder to attribute and more correlated across policyholders than traditional diversification assumptions account for.

All four papers point back to CSA's AI Controls Matrix (AICM) v1.1, whose identity, supply chain, and risk management domains provide a practical checklist organizations can use to build the AI inventories and governance documentation that underwriters are beginning to require at renewal. The AICM is available at cloudsecurityalliance.org/artifacts/ai-controls-matrix-v1-1 [12].

References

[1] fintech.global. "Undeclared AI is becoming cyber insurance's blind spot." fintech.global, September 16, 2026.

[2] Logistics Viewpoints (syndicated). "The Next Supply Chain Concentration Risk May Be the AI Model Itself." Logistics Viewpoints, September 16, 2026.

[3] Munich Re. "Cyber insurance: Risks and trends 2026." Munich Re, 2026.

[4] Verizon. "2026 Data Breach Investigations Report." Verizon, May 2026.

[5] Infosecurity Magazine. "Gartner: 40% of Firms to Be Hit By Shadow AI Security Incidents." Infosecurity Magazine, citing Gartner, November 2025.

[6] Capgemini. "World Property and Casualty Insurance Report 2026." Capgemini, May 2026.

[7] IBM. "IBM Study: One in Four Malicious Breaches are AI-Enabled, Costing Companies $6 Million on Average." IBM Newsroom, July 29, 2026.

[8] KYND. "The Wild West of AI risk: key takeaways from our first Cyber Drop Live." KYND, 2026.

[9] Cloud Security Alliance AI Safety Initiative. "The Invisible Enterprise: Shadow AI and the Ungoverned Frontier." Cloud Security Alliance, April 2, 2026.

[10] Cloud Security Alliance AI Safety Initiative. "The ChatGPT Outage Pattern: Concentration Risk in Practice." Cloud Security Alliance, July 28, 2026.

[11] Cloud Security Alliance AI Safety Initiative. "Attributing AI Attacks: When Cyber Coverage Becomes Conditional." Cloud Security Alliance, April 10, 2026.

[12] Cloud Security Alliance. "AI Controls Matrix (AICM) v1.1." Cloud Security Alliance.

[13] Cloud Security Alliance AI Safety Initiative. "Foundation Model Concentration: The Uninsurable AI Risk." Cloud Security Alliance, July 6, 2026.