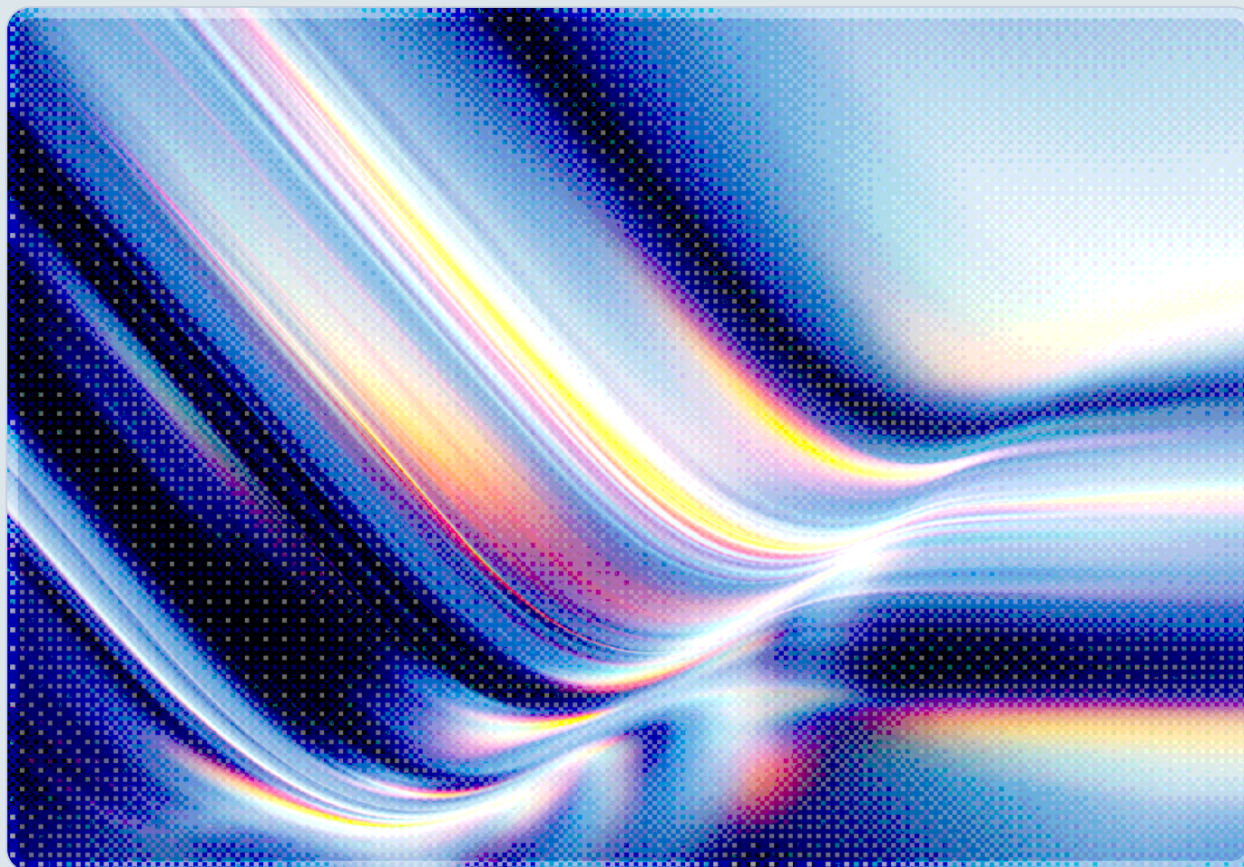


The AI Insurability Gap: Why Insurers Can't Price What They Can't See

2026-09-30

 AI-assisted Rapid Research



CSAI

cloud
CSA security
alliance®

© 2026 Cloud Security Alliance. Some rights reserved.

You may download, store, display, view, print, redistribute, and link to this document in its original, unmodified form, provided that attribution to the Cloud Security Alliance is maintained and all trademark and copyright notices remain intact.

This document may not be modified or altered. You may quote portions of the document as permitted by the Fair Use provisions of the United States Copyright Act, provided that attribution is given to the Cloud Security Alliance.

This document may be shared on professional and social media platforms in its original form with attribution.

This document was generated with AI assistance and has not undergone official CSA review and approval processes.

Key Takeaways

- Insurers evaluate emerging risks against a set of classical insurability criteria – predictable frequency, boundable severity, independence across policyholders, and freedom from severe information asymmetry – and generative AI currently strains or fails most of them, which is why carriers are responding with exclusions rather than affirmative coverage [1][4].
- The core problem is not that AI risk is merely hard to price; it is a category economists call Knightian uncertainty, where insurers lack the historical loss distribution needed to set any actuarially defensible rate at all, as opposed to ordinary risk, where a probability distribution exists even if losses are large [5].
- The Insurance Services Office's January 2026 generative AI exclusion endorsements for commercial general liability policies have been filed or adopted by more than 60 U.S. property and casualty groups, while AI-related litigation grew 978% between 2021 and 2025, illustrating how quickly the market is moving to withdraw rather than price this exposure [1][6].
- Affirmative AI-specific coverage exists but remains thin relative to the exposure it is meant to cover – OpenAI's approximately \$300 million in emerging-risk coverage sits well below the multibillion-dollar litigation exposure already in motion against foundation model providers, and specialty carriers such as CFC, Armilla AI, and Testudo Global are still building the loss history needed to underwrite AI risk at scale [1].
- Undeclared or "shadow" AI usage compounds the pricing problem by defeating the information insurers would need even if a workable actuarial model existed: with 45% of employees regularly using AI on corporate devices and two-thirds of that usage running through personal accounts invisible to IT, underwriters frequently cannot determine what AI a policyholder actually runs, let alone how it behaves under stress [2][3].

Background

Cyber and liability insurance have always depended on a foundational bargain: the insurer collects a premium calibrated against a body of historical loss data, and in exchange the policyholder transfers a risk whose frequency and severity are, at least approximately, knowable in advance. That bargain has underwritten the modern insurance industry's approach to everything from fire risk to data breaches, and it took cyber insurance roughly two decades to mature from a niche product into a mainstream corporate

expense as loss data accumulated and actuaries learned to model ransomware, business email compromise, and third-party breach exposure with reasonable confidence [1]. Generative AI has arrived on a compressed timeline that gives insurers none of that runway, and the industry's response – filing exclusions faster than it develops affirmative products – is consistent with a market that does not yet have enough information to price the risk, rather than one that has priced it and found it acceptable.

CSIS analyst Gregory C. Allen argued in a September 4, 2026 analysis that insurance has become, in effect, the most consequential de facto regulator of U.S. AI deployment, not through rulemaking but through the accumulated effect of thousands of individual underwriting decisions to exclude AI-related harms from standard commercial policies [1]. The data support the scale of that shift. More than 60 property and casualty insurance groups have filed to adopt the Insurance Services Office's new generative AI exclusion endorsements – CG 40 47, which excludes AI-related bodily injury, property damage, and personal and advertising injury claims outright, alongside the narrower CG 40 48 and CG 35 08 forms – since ISO introduced them for commercial general liability policies in January 2026 [1][6]. At least one major carrier, W.R. Berkley, has gone further, proposing to exclude coverage for "any actual or alleged use" of AI technology across its directors and officers, errors and omissions, and fiduciary liability lines [1]. Unlike earlier waves of cyber exclusions from general liability policies, which coincided with the availability of standalone cyber insurance as a replacement product, this exclusion wave is not yet being matched by a mature affirmative AI insurance market for policyholders to fall back on [1].

The regulatory backdrop is shifting in parallel but has not caught up to the underwriting retreat. More than 80% of state insurance commissioner requests to approve carrier filings excluding AI-related damages were granted as of April 2026, according to CSIS's review of state filings, even as the National Association of Insurance Commissioners' Model AI Bulletin – adopted by a growing number of U.S. states – pushes in the opposite direction by asking insurers to govern their own AI use more rigorously [1]. Demand has not slackened to match the retreat in supply: roughly 90% of corporate insurance decision-makers surveyed across major markets say they need AI-tailored coverage, according to the Geneva Association's 2025 global survey of 600 business insurance buyers, a gap between what the market wants and what it is willing to sell that is consistent with something structural, not merely cyclical, happening in AI underwriting – though confirming that will require watching whether the gap persists across underwriting cycles [1][4].

Security Analysis

One framework for understanding why insurers are retreating rather than pricing comes from actuary Karl Berliner's classical insurability criteria, a set of tests the insurance industry applies to emerging risk categories. Applying that framework to generative AI, researchers writing in *The Actuary Magazine* found

that AI risk fails outright on several criteria and is under strain on most of the rest [4]. On predictability, the creative and non-deterministic nature of generative AI outputs means losses can arise from unique interactions between a model, a prompt, and a business context that has no close historical analog, denying actuaries the repeated, similar loss events that make frequency modeling possible. On severity, potential losses from a single AI failure – a hallucinated compliance determination propagated across thousands of customer interactions, or a foundation model flaw affecting every enterprise built on it – can be large enough to threaten an insurer's capital base rather than fitting the bounded, per-incident loss pattern conventional lines assume. On information asymmetry, insurers generally cannot see how a policyholder built, validated, or constrained the AI systems it depends on, which is precisely the moral hazard and adverse selection problem insurability criteria are designed to flag [4].

Reinsurer Gen Re's August 2026 analysis distinguishing risk from uncertainty – a distinction economist Frank Knight drew in 1921 – sharpens why this differs from ordinary hard-to-price risk. Risk describes outcomes with a knowable probability distribution – even a severe outcome, like a major hurricane, can be rated because decades of data describe its frequency and range of severity. Uncertainty describes outcomes for which no such distribution exists at all, which Gen Re illustrates with the "frosted urn" problem: an actuary asked to price draws from an urn of unknown composition cannot average their way to a defensible rate no matter how sophisticated the model, because there is no distribution to average [5]. A widespread failure or compromise at a foundation model that many downstream systems depend on is, in this framing, uncertainty rather than risk: there is no accumulated history of correlated AI-model failures across thousands of enterprises to model, unlike the accumulated history of, say, regional windstorms. Gen Re's response is instructive for what it implies about where the market is headed: rather than trying to compute a rate for tail exposure that cannot be modeled, insurers are increasingly using sub-limits, widespread-event triggers, and exclusions to cap the uncertain portion of the loss rather than price it, the same contract-design logic behind Lloyd's Market Bulletins Y5381 and Y5433 requiring state-backed cyberattack exclusions [5].

Undeclared AI usage adds a second, compounding failure specifically on the information-asymmetry criterion. Verizon's 2026 Data Breach Investigations Report found that regular employee use of AI tools on corporate devices has grown to approximately 45% of the workforce, roughly quadrupling in a year, and that two-thirds of that usage occurs through personal accounts that never touch an organization's approved technology list [2][3]. KYND, a cyber risk analytics firm, framed the resulting underwriting problem during a September 2026 industry briefing: businesses are wiring undeclared AI into hiring, customer service, claims handling, and dozens of other workflows faster than insurers can ask about it, and even where an incident is clearly AI-related, determining which specific tool or model was involved is often difficult after the fact because the usage was never documented anywhere an underwriter or claims adjuster could find it [2][3]. Capgemini's 2026 World Property and Casualty Insurance Report found that 42% of insurers themselves have not measured the outcomes of their own AI initiatives, a reminder that the institutions being asked to price undeclared AI exposure in their policyholders'

operations are, in a meaningful share of cases, still working out how to govern AI in their own [2]. Where an AI-related loss can be clearly identified, IBM's 2026 Cost of a Data Breach Report found it carries a real severity premium: breaches with an AI-enabled element, chiefly deepfake-assisted impersonation and AI-generated malware, averaged \$6 million, about \$1 million above the global average across all breach types [3].

Litigation growth is compounding the pricing problem from the loss-cost side even as visibility problems compound it from the information side. Insurance broker Gallagher documented a 978% increase in generative AI-related lawsuits between 2021 and 2025, with 137% year-over-year growth in the most recent period, concentrated in patent infringement, copyright infringement, and privacy-related personal injury claims [1][6]. That growth curve is exactly the kind of rapidly shifting, thin-history loss environment actuaries are least equipped to extrapolate from, since a handful of years of steeply rising claims provide little basis for projecting where frequency and severity will stabilize.

Recommendations

Immediate Actions

Risk and legal teams should request each carrier's written position – expressly covered, expressly excluded, or silent – on how AI-instrumented losses are treated under current general liability, cyber, and technology errors and omissions wording well before the next renewal, rather than assuming prior-year language still applies given how quickly ISO's CG 40 47, CG 40 48, and CG 35 08 endorsements have spread across the market [1][6]. Organizations should also build a defensible AI usage inventory ahead of renewal, since underwriters increasingly treat undocumented AI exposure as an unrateable unknown rather than a covered risk; this inventory needs to capture shadow AI usage discovered through browser telemetry and SaaS-access reviews, not just self-reported lists from IT [2][3].

Short-Term Mitigations

Where affirmative AI coverage is unavailable or prohibitively priced, organizations should treat contract-based risk controls – indemnification and audit provisions with AI vendors, sub-limits on internally self-insured AI exposure, and explicit incident-response playbooks for AI-instrumented losses – as a substitute for insurance transfer that is not yet reliably available, mirroring the sub-limit and trigger-based approach reinsurers are using to manage their own uninsurable tail exposure [5]. Security teams evaluating specialty AI insurance products from carriers such as CFC, Munich Re, Armilla AI, or Testudo

Global should weigh coverage limits against realistic worst-case exposure; OpenAI's roughly \$300 million in emerging AI risk coverage against multibillion-dollar litigation exposure already in flight is a useful benchmark for how far current market capacity can lag actual risk [1].

Strategic Considerations

Boards and risk committees should treat the insurability gap as a likely structural, multi-year market condition rather than a temporary underwriting overreaction, given the Knightian-uncertainty and criteria-failure analysis above. Closing that gap will require the same ingredients that matured cyber insurance over the prior two decades: shared incident data, standardized loss taxonomies, and enough claims history for actuaries to move AI risk from Knightian uncertainty back toward priceable risk [1][5]. Given how vendor-concentration and cloud-dependency disclosures followed a similar path, organizations with material AI exposure should prepare for AI usage disclosure to become a standard renewal requirement, and should build the internal governance and documentation capacity to meet that requirement before it is contractually mandatory rather than after [2][4].

CSA Resource Alignment

This note's account of why insurers cannot price undeclared AI exposure connects directly to CSA's whitepaper "[The Invisible Enterprise: Shadow AI and the Ungoverned Frontier](#)," whose finding that the large majority of AI tools in enterprise environments operate outside IT control is the underlying condition that drives the information-asymmetry failure central to the Berliner insurability analysis above. That paper's 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 usage inventories that underwriters are increasingly asking for at renewal.

CSA's research note "[The AI Risk Measurement Gap](#)" makes a closely related argument from the measurement-infrastructure side of the same problem: AI risk remains difficult to price not only because insurers lack a historical loss distribution but because the industry as a whole lacks the measurement infrastructure needed to translate AI system behavior into data actuaries can use, a gap that maps onto the same Knightian-uncertainty and information-asymmetry failures this note traces through Berliner's insurability criteria.

CSA's research note "[Attributing AI Attacks: When Cyber Coverage Becomes Conditional](#)" extends this note's insurability analysis to the claims side of the same problem, showing how AI-generated malware's ability to mimic other threat actors undermines the forensic attribution that war exclusions and coverage determinations depend on – a further way in which AI erodes the "clear causation" criterion that classical

insurability analysis assumes. CSA's research note "[The ChatGPT Outage Pattern: Concentration Risk in Practice](#)" complements the severity-and-independence analysis here by documenting how dependence on a small number of foundation model providers converts an individual outage into a correlated loss event across many policyholders simultaneously, the accumulation dynamic that most directly explains why AI risk currently fails Berliner's maximum-possible-loss criterion.

All four papers point back to CSA's AI Controls Matrix (AICM) v1.1, available at cloudsecurityalliance.org/artifacts/ai-controls-matrix-v1-1, whose supply chain, risk management, and governance domains provide a practical baseline organizations can map their own AI inventory and controls against – the documentation increasingly necessary to make an AI risk profile legible enough for an underwriter to price at all.

References

- [1] Gregory C. Allen. CSIS. "[The Insurance Industry's Retreat from AI Threatens to Slow Innovation and Adoption](#)." Center for Strategic and International Studies, September 4, 2026.
- [2] fintech.global. "[Undeclared AI is becoming cyber insurance's blind spot](#)." fintech.global, September 16, 2026.
- [3] fintech.global. "[Undeclared AI is insurance's biggest blind spot](#)." fintech.global, September 25, 2026.
- [4] Martin Eling, Ruo Jia, and Tianyang Wang. "[Insights: AI Insurability](#)." The Actuary Magazine, October 2025.
- [5] Gen Re. "[Risk You Can Price, Uncertainty You Cannot: Knight's Distinction and the AI and Cyber Tails](#)." Gen Re, August 2026.
- [6] Claims Journal. "[Insurer Interest in AI Exclusions Growing as Risk Becomes Omnipresent](#)." Claims Journal, July 20, 2026.
- [7] Cloud Security Alliance AI Safety Initiative. "[The Invisible Enterprise: Shadow AI and the Ungoverned Frontier](#)." Cloud Security Alliance, April 2, 2026.
- [8] Cloud Security Alliance AI Safety Initiative. "[The AI Risk Measurement Gap](#)." Cloud Security Alliance, June 10, 2026.
- [9] Cloud Security Alliance AI Safety Initiative. "[Attributing AI Attacks: When Cyber Coverage Becomes Conditional](#)." Cloud Security Alliance, April 10, 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 Controls Matrix \(AICM\) v1.1](#)." Cloud Security Alliance.