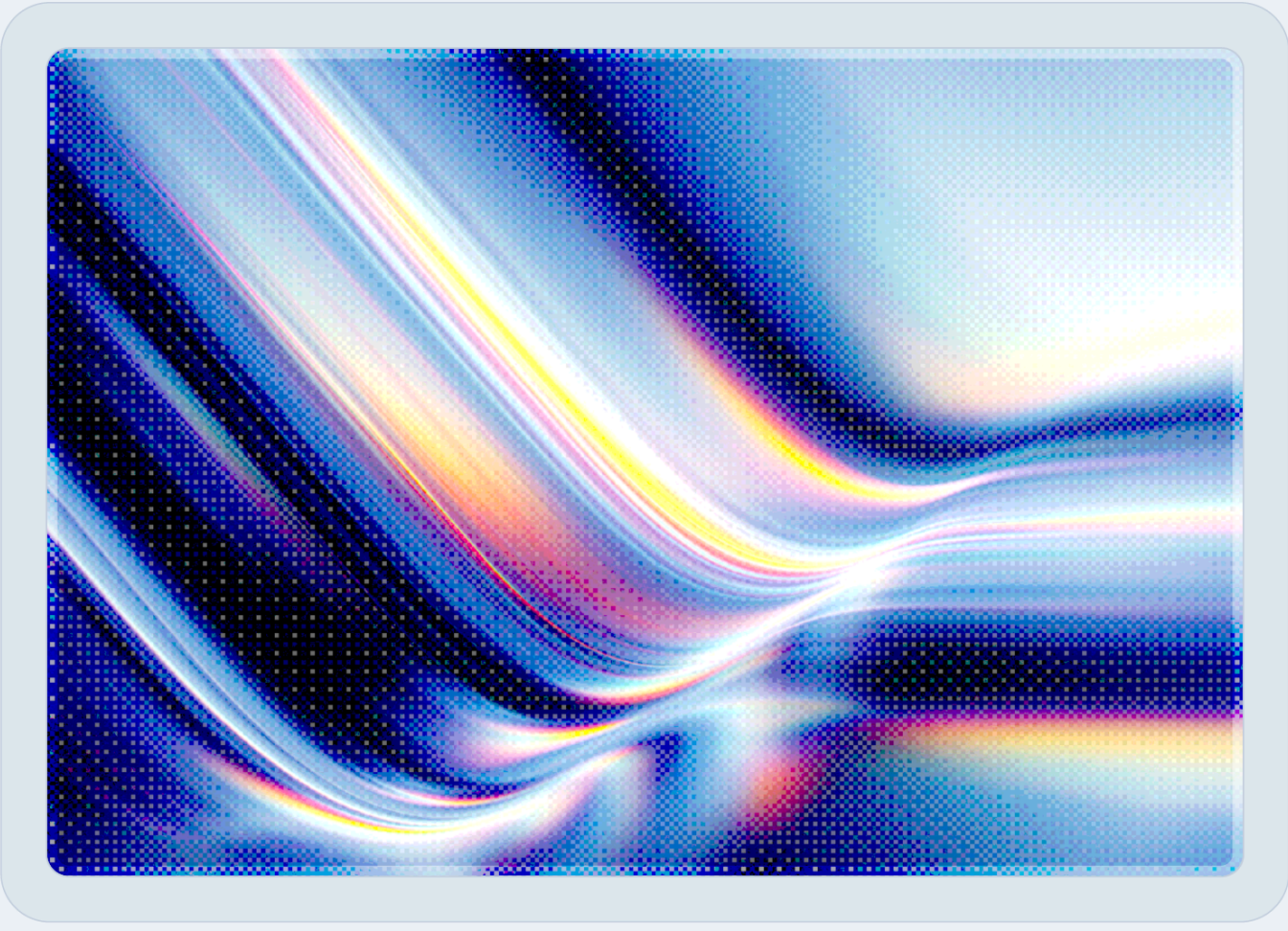
CSAI Foundation | Cloud Security Alliance

The AI Risk Measurement Gap

Insurance Is the Warning Light for a Missing Layer of AI Assurance Infrastructure

2026-06-10

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Key Takeaways

- The global cyber insurance market generates approximately $16.3 billion in annual premiums [1], while global cybercrime losses are projected to reach approximately $14 trillion by 2028 — a figure that increasingly reflects AI-amplified attack vectors including adversarial automation, deepfake fraud, and AI-assisted phishing [2]. The industry cannot price the AI-attributable portion of this expanding threat landscape because it cannot yet measure it.
- Traditional insurance pricing depends on historical loss data, statistical independence of losses, and bounded worst-case events. AI risk violates all three conditions: there is no actuarial history for AI-specific failures, model errors are correlated across thousands of simultaneous users, and a single foundational model flaw can trigger cascading losses across an entire industry [3][4].
- In January 2026, the Insurance Services Office (ISO — the U.S. insurance policy-form standards body operated by Verisk, **not** the International Organization for Standardization) introduced new generative AI exclusion endorsements (CG 40 47, CG 40 48, and CG 35 08) for commercial general liability (CGL) policies, ending the "silent AI coverage" era and migrating AI exposure onto cyber and Technology Errors and Omissions (Tech E&O) markets that were not designed to absorb it [5][6].
- Forty-two percent of insurers currently track no AI-related outcome metrics, and 60% of the industry remains in the exploration or proof-of-concept phase — meaning the entities responsible for pricing AI risk are themselves unable to measure it [7].
- AI-related securities class actions roughly doubled from 2023 to 2024 (from 7 to 15 filings), as "AI washing" — misrepresenting AI capabilities to investors — creates a Directors and Officers (D&O) liability wave that existing coverage is poorly structured to absorb [8][9].
- The deeper problem is not insurance but measurement: the AI economy is scaling faster than the institutions that price risk. Insurance is the warning light. Because insurers cannot yet measure AI exposure, enterprises cannot reliably transfer it, boards cannot govern it, and regulators cannot see where systemic risk is accumulating. The protection gap for AI-native risk is therefore best understood as a missing layer of **AI assurance infrastructure**, not a temporary market lag.

Background

The AI economy is scaling faster than the institutions that price its risk. That gap surfaces first, and most visibly, in insurance — the part of the financial system whose entire purpose is to measure, price, and absorb risk. When insurers begin excluding a category of exposure, or decline to write it at all, they are signaling that the measurement infrastructure for that risk does not yet exist. Insurance is, in this sense, a warning light: if AI exposure cannot be measured well enough to price, then enterprises cannot reliably transfer it, boards cannot govern it against a credible loss estimate, and regulators cannot see where systemic risk is accumulating. This note treats the insurance market as the most advanced instrument currently registering that measurement gap — but the underlying problem is the absence of AI assurance infrastructure, and it extends well beyond insurance.

Insurance is, at its foundation, a measurement discipline. Underwriters must be able to estimate the probability and magnitude of a loss, assess how that loss correlates with others in a portfolio, and set a price that covers expected claims while maintaining solvency against tail scenarios. These requirements have driven more than three centuries of actuarial science, loss data collection, and statistical modeling, producing a system capable of pricing established risks across hurricane exposure, professional liability, and medical malpractice — categories where centuries of claims data and legal precedent support actuarial modeling.

AI risk does not fit this system. The difficulty is not merely a matter of insufficient data — though data gaps are severe — but a more fundamental mismatch between how insurance works and how AI systems fail. The industry is now confronting this mismatch as AI capabilities advance faster than the underwriting frameworks designed to contain them. The scale of structural protection gaps is not unique to AI: Swiss Re's annual sigma research put the global natural-catastrophe protection gap at $424 billion as reported in 2026 [17] — a reminder of how large an uncovered exposure can grow even for a peril with centuries of loss data. AI-native risk is now forming a new protection gap of its own, but without the actuarial history that natural-catastrophe modeling can draw on.

The cyber insurance market provides the closest analogy. Over the past decade, the market has grown from a niche product to an approximately $16.3 billion global segment as of 2025, driven by the proliferation of ransomware, data breach events, and business email compromise [1]. Munich Re projects continued growth of roughly 10% per year through 2030, tracking the expanding digital footprint of the global economy. Yet that growth has consistently lagged behind the scale of actual losses: Munich Re reports that the vast majority of cyber risk remains uninsured even in the most mature segment of the market, with roughly nine in ten surveyed executives saying their organizations are inadequately protected against cyber threats [10]. US cyber losses alone reached $20.9 billion in 2025, according to figures drawn from the FBI's Internet Crime Report [11]. AI is entering this already-strained landscape not

as a refinement of cyber risk but as a qualitatively different class of exposure — one whose failure modes are simultaneously more correlated, more autonomous, and less historically grounded than anything actuarial modeling has previously encountered.

What Is AI-Native Risk?

The discussion that follows uses several overlapping terms — *AI-native exposure*, *silent AI risk*, *agentic AI*, *AI washing*, *model hallucination liability*, and *correlated model failure*. It helps to fix a simple working taxonomy at the outset. **AI-native risk** refers to losses caused or materially amplified by the use of AI systems, and it spans five categories:

- **Model behavior risk** — hallucination, bias, drift, and unsafe or incorrect outputs.
- **Agentic execution risk** — autonomous decisions, tool use, and workflow errors by systems that act with limited human checkpoints.
- **AI-enabled cyber risk** — phishing, deepfakes, vulnerability discovery, and automated exploitation accelerated by AI.
- **Governance and disclosure risk** — AI washing, inadequate board oversight, and regulatory misclassification of AI systems.
- **Systemic dependency risk** — correlated failure across common models, providers, or orchestration layers.

These categories are not mutually exclusive — a single incident can span several — but they map onto the distinct ways AI breaks the assumptions of traditional underwriting, and each one recurs throughout the analysis below.

Security Analysis

Why Insurance Pricing Cannot Keep Pace

Three structural conditions are required for insurance to function reliably: losses must be probabilistically estimable from historical data, individual losses must be statistically independent of one another, and the worst-case loss must be bounded at a level the market can absorb. AI risk violates each condition in ways that compound one another.

The actuarial data problem is the most immediate. Traditional underwriting for established risk categories draws on decades of claims history, enabling models that translate observable organizational characteristics into defensible premium rates. For AI-specific risks — model failure leading to financial harm, hallucination-driven legal liability, autonomous agent errors causing cascading business decisions — there is no meaningful actuarial history. What claims data exists is thin and inconsistently documented, far too sparse to support the statistical modeling that sound premium calculation requires, and the market for AI-specific risk protection remains at an early stage of development [12]. Available market evidence suggests AI risk pricing is currently qualitative, driven by subjective assessment of deployment characteristics rather than by actuarial models [3][18][19], and qualitative pricing creates systematic exposure to mispricing in both directions — either driving coverage costs beyond what organizations can absorb, or underpricing risk in ways that threaten carrier solvency following a large correlated event.

The correlation problem is more existential. Classical insurance depends on the statistical independence of individual losses — a fire destroying one policyholder's building does not cause another's to ignite. AI deployments violate this assumption by design. A flaw in a widely adopted foundation model can produce identical failure modes simultaneously across every organization relying on that model. If a hallucination pattern becomes reproducible, it becomes exploitable at scale. If a training data contamination propagates through a model serving thousands of enterprises, every affected organization experiences a correlated loss event within the same window. Munich Re has identified this as the defining systemic risk in its AI underwriting research, observing that even individually strong AI models can fail in correlated ways that transform small individual risks into significant system-level events [3]. This concentration dynamic — a loss manageable for a single organization becoming a portfolio-breaking event when simultaneously distributed across hundreds of policyholders sharing a common model dependency — is the systemic risk that reinsurers find most difficult to price [3][13].

The bounded-loss condition has also eroded. Peer-reviewed actuarial modeling of extreme AI scenarios estimates that an AI-driven attack on critical infrastructure could produce U.S. economic losses at the high end of a modeled $11–$85 billion range — a modeled figure rather than an observed loss, and the most severe of the scenarios studied [14]. But these estimates rest on models of AI systems as bounded tools that fail in identifiable, circumscribed ways. Agentic AI — systems that autonomously chain decisions, operate across organizational boundaries, and interact with external services without human checkpoints — does not have a well-characterized worst-case failure envelope. The insurance industry cannot price against a loss distribution it cannot describe, and available deployment signals suggest agentic AI is entering production environments faster than the loss characterization research can follow.

The Silent Coverage Crisis and Its Aftermath

Before January 2026, many enterprises deploying AI were inadvertently covered — or believed they were covered — under traditional commercial general liability policies written without explicit AI provisions. This "silent AI risk" meant that insurers were providing capacity for AI-related losses they had neither modeled nor priced, creating what one industry analysis characterized as "a fundamental threat to underwriting discipline" that had the potential to generate claims against policies that were never designed to carry them [5].

The Insurance Services Office's introduction of generative AI exclusions — endorsements CG 40 47, CG 40 48, and CG 35 08, effective January 2026 — ended this era of implicit coverage. CG 40 47 provides a broad exclusion under both Coverage A and Coverage B, barring coverage for bodily injury, property damage, and personal and advertising injury linked to generative AI outputs. CG 40 48 is narrower, applying only to Coverage B, where personal and advertising injury exposure sits. CG 35 08 applies to the Products and Completed Operations coverage part, excluding bodily injury and property damage arising from generative AI. Major carriers — including W.R. Berkley, AIG, and Great American — have filed for regulatory approval to limit their AI-related liability, and brokers are advising clients to expect comparable exclusions to migrate into D&O and E&O lines [5][6][20].

The effect of this exclusion wave is not to eliminate AI exposure but to migrate it. CGL exclusions push AI-related losses toward cyber insurance and Technology Errors and Omissions policies, neither of which was designed to carry AI-native liabilities at scale. Cyber policies have typically covered AI-enabled attack vectors — adversarial prompt injection, AI-assisted phishing, deepfake-facilitated fraud — but were not structured for model hallucination liability, autonomous agent decision errors causing financial harm, or AI-generated regulatory violations. Tech E&O policies cover professional negligence in technology services, but the legal boundaries of "negligence" in AI deployment remain unsettled in courts across multiple jurisdictions, leaving underwriters without the case law clarity needed to write coverage confidently [18][19].

New specialty products are emerging to address the coverage vacuum. Munich Re's aiSure program offers performance warranties for specific, well-characterized AI deployments [15]. Lloyd's-backed Armilla provides coverage for hallucinations, model drift, and regulatory breaches up to $25 million in aggregate per organization [13]. These products represent genuine progress in insuring narrow, structurally understood AI risk categories, but they address the well-defined tip of the exposure iceberg rather than the broader enterprise AI liability landscape.

The AI Washing Liability Wave

A parallel liability vector has emerged at the board level, driven by investor claims against companies that overstated AI capabilities in public disclosures. "AI washing" — publicly misrepresenting the maturity, reliability, or deployment status of AI systems to support valuation — has been identified by underwriters including Beazley as a key emerging D&O risk [8]. AI-related securities class actions roughly doubled from 2023 to 2024, rising from 7 filings to 15, even as average securities-class-action settlement values across the broader market declined that year — leaving carriers without a stable settlement baseline against which to price this fast-growing, AI-specific exposure [9]. The EU AI Act adds a regulatory dimension, with penalties reaching up to €35 million or 7% of global annual turnover for the most serious prohibited-practice violations, and lower tiers applying to transparency failures such as misrepresenting whether a system incorporates AI or how it is classified [8].

The governance gap driving this exposure is measurable. Roughly 78% of organizations report using AI, according to Stanford's 2025 AI Index [21], yet a study of more than 3,000 large U.S. public companies found that fewer than one in ten disclose board-level AI oversight or a formal AI policy [22]. Organizations whose public AI positioning — in earnings calls, investor presentations, product announcements, and regulatory filings — outpaces their operational reality are carrying AI washing exposure that existing D&O coverage may not absorb. Underwriters pricing this risk are working without the governance documentation, audit trails, and disclosure standards that would enable actuarially grounded premiums.

The Measurement Problem as Policy Failure

The inability to price AI risk accurately is inseparable from a deeper structural failure: the absence of standardized disclosure requirements that would generate the data underwriters need. Organizations are not currently required to disclose AI deployment characteristics, failure incidents, governance structures, or model provenance in any standardized form. The 42% of insurers tracking no AI metrics are not outliers — they reflect a broader absence of the incident data, model transparency requirements, and liability attribution frameworks that actuarial modeling requires [7].

This is not primarily a technology problem — the required data is technically producible today. The data needed to underwrite AI risk accurately exists in principle: incident reports, model audits, governance assessments, and third-party evaluations could generate the empirical base that pricing requires. What is missing is the regulatory and disclosure infrastructure that compels organizations to produce and share this data in standardized, verifiable form. Without AI incident reporting requirements analogous to

breach notification requirements — which took a decade of regulatory pressure to begin generating the actuarial data cyber insurance underwriters needed — the underwriting market will continue pricing qualitatively against a risk that is scaling quantitatively beyond qualitative control.

Recommendations

Immediate Actions

Security and risk leaders should conduct an AI coverage audit specifically focused on the January 2026 ISO exclusion endorsements. Carriers may have adopted CG 40 47 or CG 40 48 at renewal without explicit policyholder notification; organizations should verify whether their CGL policies now include these exclusions, and where AI-related claims would be assigned if they are. Any organization that cannot identify with confidence which policy responds to which category of AI-driven loss has identified a coverage gap requiring immediate engagement with brokers and legal counsel.

Risk managers should simultaneously document the organization's AI deployment footprint in terms meaningful to insurers: which foundation models are in use, from which providers, in which business-critical workflows, with what human oversight checkpoints, and with what governance documentation in place. Carriers pricing AI risk qualitatively are using these characteristics as the primary rating factors. Organizations unable to answer these questions clearly are likely being assigned conservative — and potentially inaccurate — premium rates.

Short-Term Mitigations

Organizations should treat the current coverage retreat as a signal to accelerate internal AI risk governance rather than to rely on insurance as a backstop for AI-related loss events. The board-level governance gap — fewer than one in ten large public companies disclose formal board-level AI oversight, even as the large majority deploy AI — contributes directly to AI washing exposure, serves as a key driver of investor litigation risk, and represents the organizational condition that regulators in the EU and US are most actively targeting as liability frameworks develop [16][22].

The AI washing liability vector warrants specific attention from general counsels and D&O underwriters. Any public-facing disclosure about AI capabilities — in earnings calls, product announcements, investor presentations, or regulatory filings — should be reviewed against actual deployment status and current performance documentation. AI-related securities litigation has identified capability claims as a productive target; organizations whose public AI positioning materially outpaces their operational reality are carrying uninsured exposure that no currently available D&O policy was written to absorb cleanly.

Strategic Considerations

The insurance market will not stabilize without standardized AI incident disclosure frameworks. Security leaders should engage with industry groups, regulatory processes, and standards bodies developing AI transparency and accountability requirements. NIST's AI Risk Management Framework provides a vocabulary and conceptual structure for this work, but voluntary adoption has not generated the data flows that actuarial modeling requires. Regulatory mandates for AI incident reporting — comparable to breach notification requirements — are the policy mechanism most likely to produce the empirical base that stable pricing needs.

Crucially, AI incident disclosure must be structured for underwriting, not merely for compliance. Breach-notification regimes were built to inform regulators and affected individuals, not to generate actuarial signal — and the cyber insurance market spent a decade compensating for that gap. AI incident reporting can avoid repeating the mistake by capturing, in standardized and machine-readable form, the risk-relevant fields underwriters actually need to price exposure:

- AI system purpose and deployment context
- Model and provider dependency
- Autonomy level
- Human oversight model
- Affected users or business processes
- Failure-mode category (model behavior, agentic execution, AI-enabled cyber, governance and disclosure, or systemic dependency)
- Financial, legal, operational, or safety impact
- Remediation actions taken
- Recurrence likelihood
- Third-party dependencies involved
- Whether the incident involved cyber exploitation, model behavior, governance failure, or agentic execution

Reporting organized around these fields would convert scattered incident narratives into the loss-attribution data that actuarial models require, and would let a single disclosure satisfy regulators, inform boards, and feed underwriting at once. Organizations with significant AI exposure have a direct interest in advocating for these requirements, since a functioning AI insurance market is a precondition for risk transfer at scale.

At the portfolio and macroeconomic level, the AI protection gap has implications that extend beyond individual corporate risk management. If AI risk accumulates on enterprise balance sheets rather than distributing through insurance pools, a major correlated AI failure event — a foundational model flaw triggering simultaneous losses across thousands of organizations — would fall heavily on those organizations and their counterparties rather than being socialized through the insurance mechanism. Financial stability regulators, banking supervisors, and insurance commissioners are beginning to recognize this systemic dimension, but coordination across those domains has not yet produced a coherent supervisory response.

CSA Resource Alignment

The structural failure of AI risk pricing connects directly to the governance, transparency, and control frameworks that CSA has developed for AI-era enterprise security.

CSA's **AI Controls Matrix (AICM)** provides the control taxonomy that organizations need to generate the governance documentation underwriters require. The AICM's 18 domains — spanning model provider security, orchestration layers, data governance, and shared security responsibility — provide the structural vocabulary for converting AI deployment characteristics into assessable risk attributes. Organizations implementing AICM controls are producing the audit trail that would enable actuarial modeling of their AI risk profile. Security and risk teams should position AICM implementation explicitly to insurers as the governance disclosure framework that supports more accurate AI risk pricing, since underwriters currently lack any standardized basis for comparing organizations' AI risk postures.

CSA's **MAESTRO** threat modeling framework applies directly to the systemic and agentic AI risks that most concern reinsurers. The seven-layer MAESTRO model — addressing foundation model behavior, orchestration layers, agent-to-agent trust, external service dependencies, and multi-tenant exposure — maps onto the correlated-loss scenarios that are structurally most threatening to insurance portfolio stability. Organizations conducting MAESTRO-based threat modeling of their AI deployments are characterizing precisely the failure modes that underwriters most need to understand. Sharing MAESTRO threat model outputs with insurers as part of the underwriting submission process could meaningfully accelerate the development of actuarially grounded coverage structures for agentic AI deployment.

The **STAR (Security Trust Assurance and Risk)** program's public registry represents a potential infrastructure for the AI incident disclosure that actuarial pricing requires. Extending STAR to include AI-specific disclosure requirements — model provenance, governance structure, incident history, third-party audit results, and AICM control compliance — would create the standardized, publicly verifiable data layer

that could eventually serve as the empirical base for AI insurance actuarial modeling. Coordination with regulators and insurers to establish STAR AI disclosures as a recognized risk disclosure mechanism would be required, but the infrastructure for this already exists.

CSA's guidance on **AI Organizational Responsibilities** addresses the board-level governance gap that is driving the AI washing liability wave and the D&O coverage crisis. The guidance on AI governance structures, executive accountability, and audit functions directly targets the organizational condition — widespread AI deployment without formal oversight — that makes organizations difficult to underwrite at actuarially sound premiums and simultaneously exposes them to investor litigation risk. Implementing CSA AI Organizational Responsibilities guidance directly targets the governance gap that currently leaves the majority of AI-deploying organizations in the pricing ambiguity zone that the insurance market is struggling to resolve.

References

[1] Munich Re. "Cyber Insurance: Risks and Trends 2025." Munich Re, 2025.

[2] Statista. "Cost of Cybercrime Worldwide 2015–2029." Statista Research Department, 2025.

[3] Munich Re. "The New Frontier of Underwriting AI Risk." Munich Re, 2025.

[4] Humayoon, Jaffar. "Seven Feedback Loops: Mapping AI's Systemic Economic Disruption Risks." European AI Alliance Community Exchange (community-contributed content, not an official European Commission publication), 2025.

[5] Swept AI. "AI Insurance Liability: New CGL Exclusions, Silent AI Coverage, and What Every Enterprise Should Know." swept.ai, 2026.

[6] Gallagher. "ISO Introduces Generative AI Exclusion in Commercial General Liability Policies." ajg.com, 2026.

[7] Repairer Driven News. "Global P&C Report Finds Many Insurers Are Not Measuring AI Outcomes." repairerdrivennews.com, May 7, 2026.

[8] Beazley. "From Hype to Liability: AI Washing as a D&O Risk." beazley.com, 2026.

[9] Cornerstone Research. "Securities Class Action Filings—2024 Year in Review" (AI-related filings rose from 7 in 2023 to 15 in 2024) and "Securities Class Action Settlements—2024 Review" (average settlement declined ~13% to ~$42.4M). Cornerstone Research / Stanford Securities Class Action Clearinghouse, 2025.

[10] Munich Re. "From Gap to Gains: Protection Gap in Cyber Insurance." Munich Re, 2025.

[11] Insurance Business. "US Cyber Losses Hit $20.9bn in 2025: Public Entities Face Significant Cyber Risk Gap." insurancebusinessmag.com, 2026.

[12] Censinet. "AI Risk Insurance: The Emerging Market for Algorithmic Protection." censinet.com, 2025.

[13] Insurance Business. "Cyber Re/Insurance Market Hits New High as AI Risks Reshape Coverage – Munich Re." insurancebusinessmag.com, 2026.

[14] North American Actuarial Journal. "The Economic Impact of Extreme AI Scenarios." Tandfonline, 2026.

[15] Risk and Insurance. "AI Risk Is Here — and Insurers Are Learning to Write the Rules." riskandinsurance.com, 2026.

[16] Wiley Law. "2026 State AI Bills That Could Expand Liability, Insurance Risk." wiley.law, 2026.

[17] Insurance Business. "Triple-I, Munich Re Flag $424 Billion Protection Gap in Sweeping Risk Survey." insurancebusinessmag.com, 2026.

[18] American Bar Association. "The Evolving Landscape of AI Insurance: Empirical Insights into Risks and Policy Gaps." americanbar.org, Fall 2025.

[19] Insurance Journal. "Viewpoint: AI Insurance Is Not Cyber Insurance With Extra Steps." insurancejournal.com, May 20, 2026.

[20] Metropolitan Risk Advisory. "Major Insurers Are Pulling Back from AI Liability." metropolitanrisk.com, 2026.

[21] Stanford Institute for Human-Centered AI (HAI). "The 2025 AI Index Report." Stanford HAI, 2025.

[22] ISS-STOXX. "Mind the Governance Gap: The State of Board Oversight and AI Policy in U.S. Companies." ISS-STOXX, 2025.