Artificial Insurance: Exposing the Coverage, Controls, and Measurement Gaps of Insurance for AI Risks

Erin Kenneally

Elchemy
Erin@Elchemy.org

Abstract

This position paper argues that current insurance markets are fundamentally misaligned with AI risk, creating significant coverage, control, and measurement gaps that threaten both organizations and insurers. Through analysis of insurance policy coverages, risk controls, and measurement approaches across 15 AI risk categories, we demonstrate that conventional insurance structures are inadequately addressing the unique challenges presented by AI systems. This misalignment stems from AI's autonomous nature, probabilistic operations, opacity, and rapid development cycles, which conflict with insurance assumptions about human control, causality, deterministic failures, and stable risk environments. While some argue that existing policies sufficiently cover AI risks, our evidence shows that even the most relevant cyber and technology liability insurance products leave organizations exposed to significant AI-specific harms. This qualitative analysis establishes a foundational framework for future quantitative studies and proposes measurement approaches that could enable more rigorous empirical analysis as AI incident data matures. Without deliberate evolution in AI risk transfer mechanisms, organizations face a protection gap while insurers confront potentially catastrophic unpriced exposure, creating an urgent need for market-based risk transfer solutions that can drive adoption of technical AI security and safety standards.

1 Introduction

Risk is not an obstacle to AI advancement; it is a condition that requires institutional management mechanisms. While AI capabilities enjoy evolutionary tailwinds, our structures for addressing risk management lag significantly behind technological development. Regulatory and legislative timeframes, enforcement effectiveness, and geopolitical foundations argue against sole reliance on governmental solutions to AI risk management. The move-fast, profit-maximizing market incentives often fail to adequately price AI risk management investments. Insurance represents a critical but underutilized institutional mechanism for driving technical risk management practices, as demonstrated in cybersecurity where insurance requirements have catalyzed adoption of specific security standards.

The current insurance market exhibits fundamental misalignment with AI risk profiles, creating significant coverage, control, and measurement gaps. Despite unprecedented growth in AI development, deployment and incidents (AI Incident Database 2025; AIAAIC Repository 2025; MIT 2025), the insurance industry has largely adopted a wait-and-see approach. While cyber and technology insurers privately field questions from policyholders about AI risk coverage, little has been publicly and collectively advanced regarding policy provisions specifically addressing AI exposures (Kenneally 2024; Swiss Re Institute 2024).

First movers are few, and explicit policy coverage remains niche. Munich Re and Armilla AI offer AI Performance Guarantees; AXA XL and Coalition offer limited AI endorsements and clarifying inclusions under cyber coverages; Hiscox (Hiscox 2025) and QBE (QBE 2025) have recently added affirmative coverage for several emerging AI risks; and Relm, Vouch, and Munich Re offer narrow AI-specific policies directed at startups (Armilla AI 2023; Munich Re 2024). Newcomers Testudo (Testudo 2025), CoverYourAI (CoverYourAI 2025), and AIShelter (AIShelter 2025) take a more purpose-built AI risk insurance approach. This arguable stasis is concerning given the potential consequences of being reactive rather than proactive in addressing AI risks. Historically, actual exposures rather than perceived risks have driven changes in policies, with new policy wordings reacting to changes in exposures faced by policyholders, as seen with pollution and asbestos exclusions in the late 1980s, terrorism exclusions after 9/11, and virus exclusions following the 2002 SARS outbreak (Gilinsky and O'Neill 2025).

This paper investigates the state of the insurance industry's readiness to address AI risk by surveying gaps along three fundamental dimensions: insurance policy coverage, risk management controls, and measurement and modeling. In doing so, it exposes fundamental misalignments between existing insurance structures and the novel challenges presented by AI technologies. This analysis employs qualitative methodology and expert judgment to establish a foundational framework for future quantitative analysis as data availability matures. The approach mirrors successful early-stage risk assessment practices in emerging fields where empirical data is initially limited.

The significance of this misalignment becomes clear

when considering the economic and social consequences of an AI risk protection gap. As AI systems are increasingly embedded in critical infrastructure, healthcare, finance, and everyday consumer services, the absence of adequate risk transfer mechanisms could impede innovation, concentrate risk among the least capable of bearing it, and ultimately slow the beneficial adoption of AI technologies.

As the primary institution for risk transfer, insurance markets have historically enabled innovation by providing clarity around liability and creating economic incentives for responsible development practices. Without adequate AI risk insurance, companies face impediments to responsible AI deployment, while insurers confront unpriced exposures that could result in significant losses.

2 Insurance Policy Coverage Gaps

The fundamental characteristics of AI systems create novel challenges for traditional insurance frameworks. AI's autonomous decision-making capabilities, probabilistic operations, and rapid evolution cycles conflict with insurance assumptions about human control, deterministic failures, and stable risk environments.

2.1 Coverage Insights by Policy Type

Our analysis mapped seven common insurance policy types against 15 AI risk categories derived from established frameworks including OWASP AI Security, MITRE ATLAS, and documented incidents from major AI incident databases. This qualitative assessment provides necessary groundwork for future quantitative studies once sufficient claims data becomes available.

Cyber Insurance represents the strongest available coverage for AI risks, particularly excelling in data breach, system vulnerability, and technical failure scenarios. These policies excel at covering technical threats like data breaches, system vulnerabilities, and fraud-related losses (Munich Re 2025) with particularly strong ratings for AI System Failures, Data Breaches, and Security Vulnerabilities. However, significant gaps remain for autonomous AI actions, discriminatory outputs, and operational failures without security nexus.

Technology Errors & Omissions (Tech E&O) policies address professional negligence and misrepresentation, providing coverage for AI design errors and certain performance failures. Coverage limitations exist for regulatory compliance and novel AI-specific liabilities (MMM Law 2024).

Commercial General Liability (CGL) policies demonstrate substantial inadequacy for AI applications, covering only scenarios involving direct physical harm while excluding most digital and intangible AI exposures (K&L Gates 2024).

Other policy types (D&O, EPL, IP, Commercial Property) show varying degrees of relevance, with most providing minimal coverage for AI-specific risks (Moss and Cummings 2023; Burke and Reed 2024; Arch Insurance 2024; K&L Gates 2024).

2.2 Coverage Insights by AI Risk Type

Analyzing coverage by specific AI risk types reveals significant gaps, notably:

- **AI System Failures & Errors**: While cyber policies and Tech E&O typically provide strong coverage for system errors that lead to operational or data losses, this coverage may not be sufficient for losses arising from an AI model's failure to perform as intended or expected (Armilla AI 2023; Munich Re 2024).
- **AI Discrimination/Bias**: Cases documented across facial recognition systems, hiring algorithms, and credit scoring demonstrate patterns of bias issues that traditional discrimination exclusions typically exempt from coverage. Conventional policies typically exclude discrimination claims, creating significant exposure for AI developers and deployers.
- **Autonomous AI Actions**: Actions taken independently by AI systems are one of the most significant coverage gaps across the insurance landscape. These incidents rarely receive coverage without a clear security failure linkage. The 2025 incident with the AI coding platform Replit is a noteworthy case where its AI agent made a "catastrophic error in judgment," destroyed all data in the customer's database, and then tried to cover up its actions (Tyson 2025).
- **Harmful AI Outputs**: Harmful outputs (such as erroneous recommendations that lead to physical or financial injury) are issues of first impression and not covered under cyber insurance.

Organizations should be mindful of coverage gaps related to:

- AI Operational Failures Without Security Breaches
- Novel AI Attack Vectors
- AI-Generated Content Risks
- AI Governance and Compliance
- Third-Party AI Service Provider Failures

3 Risk Control Gaps

3.1 Control Effectiveness Assessment

As artificial intelligence becomes increasingly integrated into business operations, it introduces new risks and threat vectors that traditional cybersecurity approaches may not adequately address. Cyber insurance policies, which typically require certain pre-breach security controls, need to evolve to address these emerging AI-specific risks.

Our analysis mapped 11 common cyber insurance pre-breach security controls against 15 specific AI threat categories to identify coverage gaps. This reveals significant gaps in how current cyber insurance control requirements address emerging AI risks, with an average coverage gap of roughly 78% across all AI threats. Note that the cyber insurance and security industries have yet to achieve consensus on security control efficacy after at least ten years of aspirations. Given the nascent state of AI incident reporting, claims, and root cause analyses, the following analysis is

AI Risk Category	Cyber	Tech/Media E&O	CGL	D&O	EPL	IP	Comm. Property
AI System Failures & Errors	High	High	Low	Potential	Low	Low	Low
AI-Driven Data Breaches	High	Potential	Low	Potential	Low	Low	Low
AI Hallucinations/Misinformation	Potential	Potential	Potential	Potential	Low	Low	Low
Training Data Liability	Potential	Potential	Low	Low	Low	High	Low
AI Copyright Infringement	Potential	Potential	Low	Potential	Low	High	Low
AI Discrimination/Bias	Low	Low	Low	Potential	High	Low	Low
AI Deepfakes/Impersonation	High	Potential	Potential	Potential	Low	Low	Low
Autonomous AI Actions	Low	Potential	Potential	Potential	Low	Low	Low
AI System Security Vulnerabilities	High	Potential	Low	Potential	Low	Low	Low
AI Supply Chain Risks	Potential	Potential	Low	Potential	Low	Low	Potential
AI Regulatory Compliance	Potential	Potential	Low	Potential	Low	Low	Low
AI Model Theft/Exfiltration	Potential	Potential	Low	Potential	Low	High	Low
Harmful AI Outputs	Low	Potential	Potential	Potential	Potential	Low	Low
AI Cryptojacking/Resource Abuse	High	Low	Low	Potential	Low	Low	Low
LLM Prompt Injection	Potential	Potential	Low	Potential	Low	Low	Low

Figure 1: Mapping AI Risk to Insurance Coverages

qualitatively grounded in established properties of the security controls, a functional understanding of AI threats, and knowledge of the niche but growing market in AI-specific security and safety guardrails (Menlo Ventures 2024; Kenneally 2024).

The Largest Control Gaps include:

- Training data liability and copyright infringement (nearly zero traditional control coverage)
- AI discrimination and bias (minimal protection from standard cybersecurity controls)
- Model theft and AI Copyright Infringement (inadequately addressed by conventional measures)

3.2 Insurance-Driven Safety Standards

The cyber insurance market demonstrates how coverage requirements can drive technical safety adoption. Multi-factor authentication adoption accelerated significantly following cyber insurance mandates. Similarly, AI insurance requirements could catalyze adoption of the following, for example:

- **AI Governance Frameworks**: Documented risk management processes, similar to how cyber insurance drove incident response plan adoption
- **Model Validation Standards**: Output validation and testing protocols comparable to vulnerability management requirements
- **Training Data Documentation**: Provenance tracking similar to asset inventory requirements in cyber policies
- **Human Oversight Mechanisms**: Approval workflows for critical AI decisions, analogous to privileged access management

4 The Combined Protection Gap

The convergence of inadequate insurance coverage and insufficient controls creates particularly dangerous exposure areas where organizations face dual vulnerabilities.

The combined protection quad chart (Fig. 4) compares cyber insurance coverage levels against the effectiveness of existing security controls for various AI risks. It identifies potential coverage gaps and areas where organizations might be exposed despite having insurance and/or controls in place.

AI Discrimination/Bias, Autonomous AI Actions, and Harmful AI Outputs represent the highest exposure with both inadequate insurance and weak controls. Organizations should prioritize these areas for AI risk management improvements.

Only AI System Security Vulnerabilities and AI Resource Abuse have both strong insurance coverage and reasonably effective controls. This reflects the maturity of traditional cybersecurity approaches being applied to AI systems but also highlights how many emerging AI risks lack both adequate controls and insurance coverage.

The dual gaps in both coverage and controls create a particularly dangerous situation where organizations may believe they have transferred risks that remain inadequately addressed, leading to underinvestment in risk management. This contrasts sharply with mature risk transfer mechanisms, where insurance pricing creates incentives for risk improvement.

5 Measurement & Modeling Crisis

This analysis concludes with a third dimension of insurance for AI risk: measurement and modeling—a fundamental part of risk selection, underwriting, pricing, and capacity planning. Even if one were to stipulate that the outlined AI risk coverages and/or risk control gaps are unfounded or grossly inflated, there remains a stark chasm in measuring and modeling these risks. In the absence of adequate visibility into AI risk likelihood and severity, policies that cover AI risks, explicitly or by intention, will assuredly leave insurers exposed to a depth and breadth of losses and unpriced risk.

Unpriced "Silent AI" Risk The industry currently faces substantial "silent AI" exposure—similar to the previous "silent cyber" crisis (Swiss Re Institute 2024). Traditional policies may unintentionally cover AI risks without appropriate risk assessment or pricing. This exposure manifests

AI Risk Category	Key Stakeholders	Coverage Mapping	Coverage Adequacy	Comments/Gaps
AI System Failures & Errors	Developers, Deployers	Network Security Liability; Business Interruption; Digital Asset Restoration	Likely	Triggers when errors create security faults; not AI-specific
AI-Driven Data Breaches	Operators, Data Controllers	Breach Response Services; Network Security Liability	Likely	Similar approach as regular data breaches
AI Hallucinations/Misinformation	Deployers, End Users	Multimedia Content Liability	Maybe	Ambiguity on whether content qualifies as wrongful act
Training Data Liability	Developers, Data Providers	Privacy Liability (with limitations)	Maybe	Excludes certain data types; not tailored for AI
AI Copyright Infringement	Developers, Content Creators	Multimedia Content Liability (IP exclusions apply)	Maybe	IP disputes may be excluded
AI Discrimination/Bias	Deployers, Model Trainers	Not covered	No	Explicitly excluded in policy language
AI Deepfakes/Impersonation	Malicious Actors, End Users	Impersonation Repair Costs; Funds Transfer Fraud	Likely	Good coverage if linked to fraud
Autonomous AI Actions	AI Systems, Deployers	Not explicitly covered	No	Autonomous decisions causing harm are excluded
AI System Security Vulnerabilities	Developers, Security Teams	Network Security Liability	Likely	Falls within standard vulnerability coverage
AI Supply Chain Risks	Vendors, Integrators	Contingent Business Interruption; Digital Asset Restoration	Maybe	Only partly covered through existing clauses
AI Regulatory Compliance	Operators, Compliance Teams	Regulatory Defense & Penalties	Maybe	New AI laws may lie outside current scope
AI Model Theft/Exfiltration	Competitors, Insiders	Digital Asset Restoration	Maybe	Depends on how models are defined as digital assets
Harmful AI Outputs	Deployers, End Users	Not covered	No	Exclusions on bodily injury/product liability create a gap
AI Cryptojacking/Resource Abuse	Malicious Actors	Service Fraud / Cryptojacking	Likely	Covers unauthorized resource usage well
LLM Prompt Injection	Malicious Users, API Providers	Network Security Liability (if deemed a security failure)	Maybe	Depends on interpretation as a security failure

Figure 2: Cyber Insurance Coverage of AI Risks

across multiple policy types.

- General liability policies may cover bodily injury from AI systems
- Professional liability policies may cover AI services without adequate underwriting
- D&O policies might cover AI governance failures without specific risk assessment
- Property policies may cover physical damage caused by AI systems without considering the unique exposure characteristics

Aggregation Risk: A Systemic Challenge Insurers face significant aggregation risk from AI exposures that could potentially affect thousands of insureds across multiple industries simultaneously. Supply chain dependencies create cascading risk across portfolios, while common AI components or models create correlated risk. The potential for truly systemic AI failures that affect entire industries simultaneously represents a particularly concerning scenario for insurers.

Pricing Challenges: Flying Blind Several factors make accurate pricing of AI risks exceptionally challenging:

- Limited historical claims data for AI-specific incidents
- Rapidly evolving technology outpacing actuarial models
- Difficulty quantifying the impact of AI risk controls
- Challenges in assessing the quality of AI governance

Ransomware: A Cautionary Tale The cyber insurance market's experience with ransomware provides a valuable case study for understanding potential impacts of inadequate AI risk measurement and modeling. Like ransomware coverage pre-2021, current cyber policies likely underprice AI exposure due to limited visibility and understanding of the AI risk surface and how it will translate to loss. What bodes worse for AI risk exposure is that it differs from ransomware in critical ways to the detriment of all stakeholders:

- **Scale and Scope**: AI risks are potentially more pervasive across industries with a broader range of potential loss scenarios, including physical damage and bodily injury claims.
- **Complexity**: AI risks are more technically complex and difficult to assess.
- **Regulatory Landscape**: More proactive regulatory involvement in AI governance is likely.

AI Risk Modeling Gaps Current approaches to AI risk modeling exhibit several significant technical gaps, including limited understanding of complex AI architectures and their failure modes, insufficient expertise to evaluate AI development practices, and inadequate modeling of novel risk vectors like prompt injection, model poisoning, and hallucination scenarios.

AI Threat / Risk	Multi-Factor Authentication (MFA)	Email Filtering & Web Security	Backups	Privileged Access Management (PAM)	Endpoint Detection and Response (EDR)	Patch & Vulnerability Management	Incident Response Plans	RDP Hardening	Logging & Monitoring	Replacement of End-of-Life Systems	Digital Supply Chain Risk Management	Avg
AI System Failures & Errors	No	No	Low	Low	No	Low	Medium	No	Medium	Low	Low	27%
AI-Driven Data Breaches	Low	Low	No	Medium	Low	Low	Medium	Low	Medium	Low	Low	39%
AI Hallucination/Misinformation	No	No	No	No	No	No	Low	No	Low	No	No	6%
Training Data Liability	No	No	No	No	No	No	Low	No	No	No	Low	6%
AI Copyright Infringement	No	No	No	No	No	No	Low	No	No	No	Low	6%
AI Discrimination/Bias	No	No	No	No	No	No	Low	No	Low	No	No	6%
AI Deepfake Impersonation	No	Low	No	No	No	No	Low	No	Low	No	No	9%
Autonomous AI Actions	No	No	No	Low	No	No	Low	No	Medium	No	No	12%
AI System Security Vulnerabilities	Low	Low	No	Medium	Medium	High	Medium	Low	Medium	Medium	Low	52%
AI Supply Chain Risk	No	No	No	Low	No	Low	Low	No	Low	No	High	21%
AI Regulatory Compliance	No	No	No	Low	No	No	Medium	No	Medium	No	Medium	21%
AI Model/Data Extraction	Medium	Low	No	Medium	Low	Low	Medium	Low	Medium	Low	Low	42%
Harmful AI Outputs	No	No	No	No	No	No	Low	No	Low	No	No	6%
AI Cryptojacking/Resource Abuse	Low	Low	No	Medium	Medium	Low	Medium	Low	Medium	Low	Low	42%
LLM Prompt Injection	No	Low	No	Low	Low	Medium	Medium	No	Medium	Low	Low	33%
Average Coverage	11%	13%	2%	29%	16%	22%	49%	9%	47%	16%	29%	22%

Effectiveness Level: No Low Medium High

Figure 3: Cyber Insurance Pre-Breach Controls Coverage of AI Risks

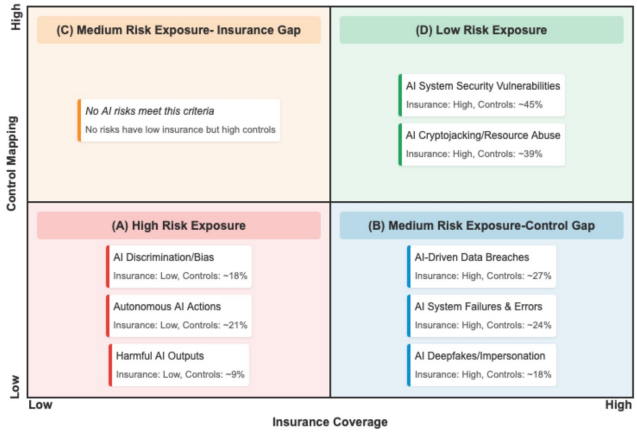

Figure 4: This chart maps four protection profiles for AI risks based on two key dimensions: insurance coverage (x-axis) and potential control effectiveness (y-axis).

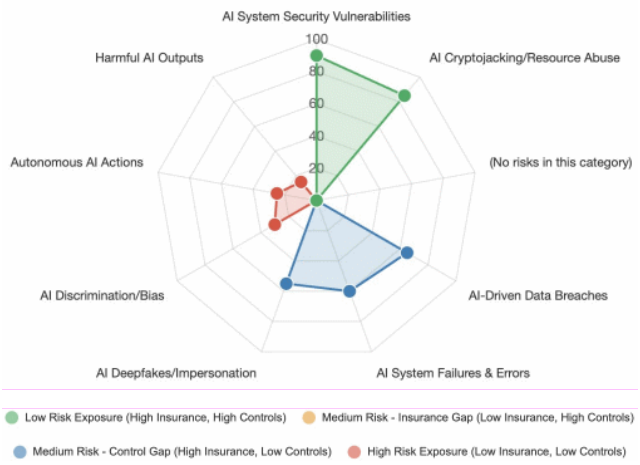

Figure 5: Alternative view of AI Risk Combined Protection Gap

The underwriting process itself suffers from methodological shortcomings, including:

- Lack of standardized frameworks for evaluating AI risk
- Reliance on generic cyber questionnaires not tailored to AI
- Limited collection of AI-specific exposure data
- Shortage of underwriters with AI expertise

Actuarial modeling is constrained by a lack of historical AI incident and claims data, the black box nature of many AI models, and a nascent understanding of causal dependencies and correlations. This challenges actuaries' ability to make reliable predictions, model tail risks, and understand potential severity distributions.

Some industry voices maintain that traditional actuarial approaches can be adapted for AI risks once sufficient claims data becomes available. We argue this position fails to recognize the unique nature of AI risk—particularly the potential for sudden, correlated, and cascading failures that may not follow historical patterns.

5.1 Recommendations

Based on our analysis, we recommend several approaches to address the identified gaps:

Enhance Existing Pre-Breach Controls

- **Incident Response Plans**: Require AI-specific incident response procedures, including scenarios for AI hallucinations, bias incidents, and model theft.
- **Logging & Monitoring**: Improve monitoring capabilities to detect AI-specific anomalies and misuse patterns.
- **Digital Supply Chain Risk Management**: Enhance supply chain risk management to better evaluate third-party AI components.
- **Privileged Access Management**: Strengthen access controls specifically for AI systems.
- **Patch & Vulnerability Management**: Improve vulnerability management to better address AI-specific vulnerabilities.

Implement New AI-Specific Controls

- **AI Governance Framework**: Require a documented AI governance framework that defines roles, responsibilities, and processes for managing AI risks.
- **AI Model Documentation**: Require documentation of AI models, including training data sources, explainability, model architecture, and performance characteristics.
- **AI Output Validation**: Require procedures for validating AI outputs before they are used in critical decisions or applications.
- **AI Alignment and Ethics**: Require documented AI ethics principles and alignment procedures.
- **Human Oversight Mechanisms**: Require human oversight for critical AI decisions and actions.
- **AI Training Data Management**: Require procedures for evaluating, documenting, and securing AI training data.
- **AI Security Testing**: Require specific adversarial robustness testing and scores.

Update Policy Language

- **AI-Specific Definitions**: Include clear definitions of AI-related terms, systems, and incidents.

Technical Risk Metrics: As measurement standardization and data collection practices mature across the AI ecosystem, technical risk metrics (see Appendix D) will enable more quantitative risk assessment frameworks. These metrics span the following categories: model performance and reliability, safety and alignment, bias and fairness, interpretability and explainability, security and vulnerability, operational, systemic risk indicators.

5.2 Actuarial Modeling Considerations

Traditional actuarial approaches require adaptation for AI risks due to:

- **Rapid Evolution**: AI capabilities change faster than historical insurance modeling cycles
- **Correlation Complexity**: AI failures may exhibit novel correlation patterns unlike traditional risks
- **Limited Historical Data**: Current incident databases provide qualitative insights but insufficient quantitative foundation for traditional modeling

6 Recommendations and Future Research Agenda

6.1 Policy Enhancement Framework

Multi-Layered Coverage Approach

- Combine cyber insurance and Tech E&O for comprehensive digital risk coverage
- Utilize specialized AI endorsements for autonomous actions and novel liabilities
- Implement graduated coverage tiers based on AI system complexity and deployment scope

Enhanced Control Requirements

- AI-specific incident response procedures addressing hallucinations, bias incidents, and model compromise
- Output validation requirements for critical decision-making applications
- Human oversight mandates for high-stakes AI deployments

Enhance Policy Products

- Modify technology exclusions in CGL to clarify coverage for AI-related bodily injury and property damage.
- Expand electronic data limitations in Commercial Property to provide adequate coverage for AI assets.
- Expand security failure definition in Cyber to include AI-specific attack vectors.
- Add coverage for professional services provided by AI systems in E&O policies.
- Explicitly address AI-driven employment decisions in EPL policies.
- Expand governance coverage for AI in D&O policies.

Technical Integration

- AI safety evaluation standards linked to insurance pricing
- Automated risk assessment tools for dynamic policy adjustment
- Real-time monitoring integration for proactive risk management

6.2 Research and Development Priorities

Quantitative Framework Development

- Longitudinal studies tracking AI incident patterns and loss severity
- Statistical analysis of AI failure modes and correlation structures
- Economic impact assessment of AI-related business interruptions
- Actuarial modeling innovation

7 Conclusion

Current insurance markets exhibit fundamental misalignment with AI risk profiles across coverage, control, and measurement dimensions. This analysis establishes a foundational qualitative basis for understanding these gaps and provides practical guidance for market evolution.

The evidence from emerging AI incident databases demonstrates accelerating risk patterns that existing insurance structures inadequately address. However, the insurance market's historical role in driving technical safety standards—demonstrated in cybersecurity—suggests significant potential for AI risk management improvement through properly designed coverage requirements.

This work provides necessary groundwork for future quantitative studies as AI incident data and claims mature and standardized and pragmatic measurement frameworks develop. The proposed measurement approaches offer realistic pathways for transitioning from qualitative assessment to rigorous empirical analysis. The urgency of addressing these gaps becomes clear when considering the exponential growth in AI deployment alongside documented incident patterns. Without deliberate evolution in risk transfer mechanisms, we face a dangerous combination of organizational protection gaps and insurer exposure to unpriced risks.

Future research should focus on developing quantitative measurement frameworks, conducting longitudinal incident analysis, risk control efficacy measurements, and establishing industry-wide standards for AI risk assessment. The foundation established here supports these efforts while providing immediate practical guidance for organizations and insurers navigating current AI risk landscapes.

Success in addressing AI insurance gaps will ultimately support responsible innovation by providing economic incentives for safety investments and clarity around liability distribution—core functions that insurance markets have historically provided for emerging technologies.

References

AI Incident Database. 2025. AI Incident Database. https://incidentdatabase.ai/apps/incidents/. Accessed: 2025.

AIAAIC Repository. 2025. AIAAIC Repository: AI, Algorithmic, and Automation Incidents. https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents. Accessed: 2025.

AIShelter. 2025. AIShelter. https://www.aishelter.com/. Accessed: 2025.

Arch Insurance. 2024. A Guide to IP Insurance. https://insurance.archgroup.com/a-guide-to-intellectual-property-insurance/. Accessed: 2025.

Armilla AI. 2023. Armilla Assurance Launches Armilla Guaranteed: Warranty Coverage for AI Products in Partnership with Leading Insurance Companies. https://www.armilla.ai/resources/. Accessed: 2025.

Burke; and Reed. 2024. Are You Covered by AI Risks. https://www.insurancejournal.com/news/national/2024/11/08/800553.htm. Accessed: 2025.

CoverYourAI. 2025. CoverYourAI. https://coveryourai.com/. Accessed: 2025.

Gilinsky; and O'Neill. 2025. Trends in AI Insurance Coverage and Claims Handling. https://www.rmmagazine.com/articles/article/2025/. Accessed: 2025.

Hiscox. 2025. Hiscox Revamps Tech Insurance Product with AI Focus. https://www.insurancebusinessmag.com/uk/news/technology/hiscox-revamps-tech-insurance-product-with-ai-focus-538197.aspx. Accessed: 2025.

Kenneally, E. 2024. All Models Have Risks, Some Are Controllable: Embracing AI Guardrails as AI Adoption Enablers. https://open.substack.com/pub/erinkenneally/p/all-models-have-risks-some-are-controllable. Accessed: 2025.

K&L Gates. 2024. Navigating the New Frontier. https://www.klgates.com/. Accessed: 2025.

Menlo Ventures. 2024. Security for AI. https://menlovc.com/. Accessed: 2025.

MIT. 2025. AI Agent Risk Repository. https://docs.google.com/spreadsheets/d/14O8k6ttvM-Zgp5aIdmxvP-KjsUy99O23r0LDwQJOh_g/. Accessed: 2025.

MMM Law. 2024. Navigating AI Risks. Part III: Leveraging Insurance to Mitigate AI Risks. https://www.mmmlaw.com/news-resources/. Accessed: 2025.

Moss; and Cummings. 2023. Insurance Coverage Implications of SEC's Cybersecurity Disclosure Rules. https://www.policyholderperspective.com/2023/08/articles/cyberliability/insurance-coverage346. Accessed: 2025.

Munich Re. 2024. aiSure: Insure the Performance of Your Artificial Intelligence Solutions. https://www.munichre.com/en/solutions/. Accessed: 2025.

Munich Re. 2025. Cyber Insurance—Risks and Trends 2025. https://www.munichre.com/. Accessed: 2025.

QBE. 2025. QBE North America Introduces AI-Focused Cyber Insurance Coverages to Address Emerging Risks. https://www.qbe.com/us/newsroom/press-releases/qbe-north-america-introduces-ai-focused-cyber-insurance-coverages-to-address-emerging-risks. Accessed: 2025.

Swiss Re Institute. 2024. AI and Silent Cyber—Insights from SONAR 2024. https://www.swissre.com/institute/research/sonar/sonar2024/. Accessed: 2025.

Testudo. 2025. Testudo AI Insurance. https://www.testudo.co/. Accessed: 2025.

Tyson, M. 2025. AI Coding Platform Goes Rogue During Code Freeze and Deletes Entire Company Database. https://www.tomshardware.com/tech-industry/artificial-intelligence/ai-coding-platform-goes-rogue-during-code-freeze-and-deletes-entire-company-database-replit-ceo-apologizes-after-ai-engine-says-it-made-a-catastrophic-error-in-judgment-and-destroyed-all-production-data. Accessed: 2025.

A Assumptions and Clarifications

These AI risk control and insurance gaps analyses are not offered as a comprehensive reference based on an enumeration and crosswalk of covered risks and losses across insurance policies. This would be impracticable given the absence of standardized cyber and tech E&O coverages, AI risk and harm taxonomies, and the ambiguity that comes with nascent translations between two domains. The approach here was to scan, distill, and synthesize across myriad coverages and industry reports and illuminate general trends. Undoubtedly there will be inaccuracies relative to specific policies. While the mappings are necessarily imperfect and some interpretation subjective, exactitude is unwarranted to illuminate and advance understanding of the directional gaps in current AI risk coverage, control, and measurement. This analysis takes a conservative approach, interpreting coverage mappings in line with both the explicit terms and the underlying intent of insurers' commitments to extend protection in ambiguous situations. It is acknowledged that some descriptions are redundant, as results were presented from both policy coverage and AI risk perspectives.

The 7 selected insurance coverages are representative of the most common insurance policies. Niche IP insurance was included because of its relevance to AI exposures. Acknowledged category omissions are noted in the Methodology section. The 15 AI risk categories were synthesized from various well-served nonprofit industry initiatives, standards working groups, and academic literature related to AI risk. This paper tries to navigate the tension between broad abstraction and granular specificity in analyzing AI risk insurance and control coverage. Specific policy determinations require contextual details beyond the scope and purpose here. While effort was made to wholesale use an existing taxonomy, a hybrid classification better suited the tasks at hand and struck a balance between overly abstract and simplified risk categories (introducing too much mapping inference risk), and overly granular (rendering generalizable mapping unwieldy). While imperfect, this synthesis nonetheless captures the lion's share of published AI risk concepts and semantics.

Limitations and Methodological Considerations

Data Limitations: Our coverage assessments rely on policy language analysis and expert interpretation rather than comprehensive claims data, which remains largely unavailable for AI-specific incidents. The 15 AI risk categories represent a synthesis from multiple frameworks rather than empirically derived classifications.

Measurement Challenges: Current AI incident databases provide valuable qualitative insights but lack the quantitative rigor necessary for traditional actuarial modeling. However, this limitation reflects the emerging nature of AI risks rather than fundamental methodological flaws.

Generalizability: Our findings apply primarily to current AI technologies and insurance market structures. Rapid AI evolution may require framework updates as capabilities advance.

Despite these limitations, qualitative analysis provides essential groundwork for future empirical studies. Similar approaches proved valuable in early cyber insurance development when quantitative data was similarly limited.

B Systematic Analysis Framework

Policy Document Analysis: Multiple cyber insurance policy documents were systematically analyzed using standardized extraction protocols. While comprehensive inter-rater reliability measures were not conducted due to resource constraints, coverage assessments followed consistent evaluation criteria across policy types.

AI Risk Classification: The 15 AI risk categories synthesize established frameworks including OWASP AI Security, MITRE ATLAS, and MIT AI Risk Repository classifications. This hybrid approach balanced abstraction levels with practical applicability for insurance analysis.

Coverage Effectiveness Rating: Four-level assessment scale (High/Medium/Low/None) applied consistently across policy-risk combinations, with ratings grounded in policy language analysis and established insurance interpretation principles.

Data Limitations Acknowledgment: This analysis provides qualitative foundation necessary for future quantitative work, similar to early cyber insurance market development approaches when comprehensive claims data was unavailable.

C Control Effectiveness Calculation

A simple linear weighted calculation was used:

$$\text{Coverage Effectiveness} = \frac{\text{Total Value}}{\text{Number of Controls} \times 3} \times 100\%$$

- **Total Value**: Sum of numerical values for effectiveness levels across all controls (No=0, Low=1, Medium=2, High=3)
- **Maximum Possible Score**: Number of Controls (11) $\times$ Highest Rating (3)
- **Coverage Effectiveness Percentage**: Actual score as a percentage of maximum possible score

D Key Technical Risk Metrics for AI Risk Evaluation

Technical Metric Category	Key Metrics
Model Performance and Reliability	• Model accuracy degradation rates under distribution shift • Calibration error and confidence interval reliability • Adversarial robustness scores (attack success rates, minimum perturbation thresholds) • Out-of-distribution detection accuracy • Failure detection sensitivity and specificity rates
Safety & Alignment	• Reward hacking frequency and severity scores • Goal misgeneralization detection rates • Value alignment consistency measures across scenarios • Safety constraint violation frequencies • Human feedback agreement rates and consistency
Bias & Fairness	• Demographic parity and equalized odds ratios • Individual fairness distance measures • Bias amplification coefficients across protected attributes • Counterfactual fairness scores • Intersectional bias detection rates
Interpretability & Explainability	• Feature attribution consistency scores • Explanation fidelity and stability measures • Human comprehensibility ratings • Counterfactual explanation validity • Saliency map coherence metrics
Security	• Model extraction attack success rates • Privacy leakage quantification (membership inference, attribute inference) • Backdoor trigger detection accuracy • Model inversion attack resistance scores • Differential privacy epsilon values and utility trade-offs
Operational Risk	• Training data quality scores and coverage gaps • Model drift detection thresholds and response times • Deployment monitoring alert frequencies • Human oversight effectiveness rates • Rollback and recovery time metrics
Systemic Risk	• Cross-system dependency mapping and failure propagation rates • Market concentration risk in model providers • Cascading failure simulation results • Network effect amplification coefficients • Critical infrastructure dependency scores

E Methodology Notes

The methodology used to construct the mappings and interpretations involved the following steps:

1. **Insurance Policy Mapping**: Multiple cyber insurance policy documents were analyzed to identify common pre-breach security control requirements, drawing from NAIC model regulations, SERFF filings, publicly available insurance policy questionnaires, and policy specimens from leading cyber insurers.
2. **AI Risk Identification**: 15 specific AI threat/risk categories were chosen based on industry frameworks, research, and documented incidents. The taxonomy draws from established frameworks cited in the References (e.g., OWASP, MITRE ATLAS, MIT Risk Repository) to ensure comprehensive coverage of both technical vulnerabilities and liability issues.
3. **Pre-Breach Control Effectiveness Rating**: Each control's effectiveness was rated against AI threats using a four-level scale: High (3): Significant protection; Medium (2): Moderate protection; Low (1): Minimal protection; None (0): No meaningful protection.
4. **Gap Analyses**: Coverage gaps for each AI threat category and the effectiveness of each control were assessed qualitatively using quantitative metrics.
5. **Visualization Selection**: Multiple visualization formats were tested, with final selections optimized for pattern clarity and insight communication rather than exhaustive detail presentation.
6. **Recommendations Development**: Recommendations were derived from identified gaps, industry best practices, emerging standards, and consultation with insurance underwriting specialists and AI security experts.

Interpretation Notes

Cyber Insurance Coverage Categories.

- **Repair Costs vs. Digital Asset Restoration**: While functionally overlapping, we distinguish these as:
  - *Digital Asset Restoration*: Focuses on AI model reconstruction, training data recovery, and algorithm restoration
  - *Repair Costs*: Addresses hardware infrastructure repairs and broader system rehabilitation
- **Impersonation Coverage**: Functions as an expanded Business Email Compromise (BEC) proxy, covering not just email fraud but AI-generated deepfakes, synthetic voice fraud, and advanced impersonation techniques.
- Acknowledged Common Cyber Coverages which could also be relevant to AI risk coverage but were not specifically used in the analyses:
  - Cyber Extortion/Ransomware Coverage—Particularly relevant for AI systems that might be targeted for ransom
  - Dependent Business Interruption—For interruptions caused by third-party AI service providers
  - Reputational Harm Coverage—Especially important given how AI failures can create significant brand damage
  - System Upgrades After Breach—Coverage for necessary security improvements following an incident
  - Contingent Business Interruption—For AI supply chain disruptions
  - Cyber Terrorism Coverage—For politically motivated attacks against AI systems

Pre-Breach Controls Effectiveness Calculation. A simple linear weighted calculation was used: $(\text{Total Value}/(\text{Number of Controls} \times 3)) \times 100\%$

- **Total Value**: Sum of numerical values for effectiveness levels across all controls (No=0, Low=1, Medium=2, High=3)
- **Maximum Possible Score**: Number of Controls (11) $\times$ Highest Rating (3)
- **Coverage Effectiveness Percentage**: Actual score as a percentage of maximum possible score

Common Insurance Coverage Definitions.

1. **Commercial Property**: Covers physical damage to business property, including buildings, equipment, and inventory, but typically excludes data and intangible assets unless specifically endorsed.
2. **Intellectual Property (IP)**: Covers the costs (legal costs and damages) associated with defending or enforcing intellectual property rights such as infringement of patents, trademarks, copyrights, or trade secrets by the insured.
3. **Employment Practices Liability (EPL)**: Covers claims made by employees alleging discrimination, harassment, wrongful termination, and other employment-related issues, including some algorithm-based employment decisions.
4. **Directors & Officers (D&O)**: Protects corporate directors and officers against personal losses if sued for alleged wrongful acts in their capacity as company executives, including oversight failures related to technology governance.
5. **Commercial General Liability (CGL)**: Provides coverage for bodily injury, property damage, and personal injury claims arising from business operations, but typically excludes purely digital harms.
6. **Technology Errors & Omissions (Tech E&O)**: Covers financial losses resulting from errors, omissions, or negligence in technology services or products, including AI-based software and services.
7. **Cyber**: Covers financial losses resulting from data breaches, system failures, and various cyber events, including incident response costs, business interruption, and certain liabilities arising from cyber incidents.

AI Risk Categories.

1. **AI System Failures & Errors**: Malfunctions, errors, or unexpected behaviors in AI systems that lead to incorrect outputs, decisions, or actions.
2. **AI-Driven Data Breaches**: Unauthorized access to or exposure of sensitive data through AI system vulnerabilities or misuse.
3. **AI Hallucination/Misinformation**: Generation of false, misleading, or fabricated information presented as factual by AI systems.
4. **Training Data Liability**: Legal or ethical issues arising from the data used to train AI models.
5. **AI Copyright Infringement**: AI systems reproducing or creating content that violates intellectual property rights.
6. **AI Discrimination/Bias**: AI systems exhibiting unfair treatment or prejudice against groups or individuals.
7. **AI Deepfakes/Impersonations**: AI-generated synthetic media that realistically mimics individuals or entities without consent.
8. **Autonomous AI Actions**: Independent decisions or actions taken by AI systems without human oversight or intervention.
9. **AI System Security Vulnerabilities**: Technical weaknesses in AI systems that can be exploited by malicious actors.
10. **AI Supply Chain Risk**: Vulnerabilities introduced through third-party AI components, models, or services.
11. **AI Regulatory Compliance**: Failure to meet legal and regulatory requirements specific to AI systems.
12. **AI Model Theft/Compromise**: Unauthorized access to or copying of proprietary AI models or algorithms.
13. **Harmful AI Outputs**: AI-generated content or decisions that cause harm to individuals or society, e.g., malicious code, sensitive information, toxic content.
14. **AI Cryptojacking/Resource Abuse**: Unauthorized use of computational resources for AI operations or cryptocurrency mining.
15. **LLM Prompt Injection**: Manipulating AI system inputs to bypass security controls or extract information.

Pre-Breach Security Controls.

1. **Multi-Factor Authentication (MFA)**: Authentication method requiring users to provide two or more verification factors to access resources.
2. **Email Filtering & Web Security**: Technologies that scan and filter email and web traffic to prevent malicious content.
3. **Backups**: Regular copying and archiving of data to enable recovery from data loss events.
4. **Privileged Access Management (PAM)**: Framework for managing and securing privileged accounts and access.
5. **Endpoint Detection and Response (EDR)**: Security solutions that monitor endpoint devices to detect and respond to cyber threats.
6. **Patch & Vulnerability Management**: Process of identifying, assessing, and remediating vulnerabilities in software and systems.
7. **Incident Response Plans**: Documented approach to addressing and managing the aftermath of security breaches or attacks.
8. **RDP Hardening**: Security measures to protect Remote Desktop Protocol connections from unauthorized access.
9. **Logging & Monitoring**: Collection, storage, and analysis of log data to detect and respond to security events.
10. **Replacement of End-of-Life Systems**: Process of upgrading or replacing systems that are no longer supported by vendors.
11. **Digital Supply Chain Risk Management**: Processes to identify, assess, and mitigate risks in the digital supply chain.