Research Report

SASHA ROMANOSKY, CELINE ROBINSON

The Insurability of Artificial Intelligence

For more information on this publication, visit **www.rand.org/t/RRA5130-1**.

About RAND

RAND is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest. To learn more about RAND, visit www.rand.org.

Research Integrity

Our research integrity is grounded in RAND's core values of quality and objectivity. Rigorous quality assurance procedures, conflict of interest screening, and transparency in funding ensure that every study is objective and nonpartisan. Learn more at www.rand.org/integrity.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Published by the RAND Corporation, Santa Monica, Calif.
© 2026 RAND Corporation
RAND® is a registered trademark.

Limited Print and Electronic Distribution Rights

This publication and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited; linking directly to its webpage on rand.org is encouraged. Permission is required from RAND to reproduce, or reuse in another form, any of its research products for commercial purposes. For information on reprint and reuse permissions, visit www.rand.org/about/publishing/permissions.

RR-A5130-1

About This Report

Businesses increasingly embed artificial intelligence (AI) in operations, products, and services, creating both new opportunities and new forms of risk. Firms are deploying AI in customer service, employment screening, coding, fraud detection, and autonomous systems. At the same time, AI-related harms are already emerging, including incorrect or misleading outputs, deepfakes, privacy violations, intellectual property disputes, fraud, product defects, and discriminatory decisions. These harms create demand for insurance but do not fit neatly within existing insurance lines.

In this report, we examine the insurability of AI using a literature review, expert interviews, and analysis of public data on AI incidents, lawsuits, enacted state laws, and admitted-market insurance filings. We find that misinformation, deepfakes, harmful content, and agentic failures dominate public incidents, while lawsuits remain concentrated in intellectual property disputes over model training but are increasingly expanding into privacy, fraud, defamation, and product liability claims. Insurers are responding unevenly, with some offering affirmative AI coverage through endorsements or stand-alone AI products and others filing broad exclusions or remaining silent.

We also discuss opportunities for AI systems to generate accumulation (correlated) AI risk, including shared model dependencies, infrastructure concentration, universal attacks, and legal shocks.

This report is intended primarily for insurers, reinsurers, brokers, and managing general agents, as well as for state insurance regulators and policymakers, including the National Association of Insurance Commissioners. The objective is to inform near-term decisions about AI coverage, underwriting, market oversight, and the management of correlated AI risk.

This work was conducted within the RAND Institute for Civil Justice and the RAND Feinberg Center for Catastrophic Risk Management and Compensation.

RAND Institute for Civil Justice

The RAND Institute for Civil Justice is dedicated to improving the civil justice system by supplying policymakers and the public with rigorous and nonpartisan research. The institute's research takes an interdisciplinary, empirical approach, and its peer-reviewed publications are disseminated widely to policymakers, practitioners in law and business, other researchers, and the public. The institute is part of RAND Education, Employment, and Infrastructure. For more information, visit www.rand.org/icj or email EEI@rand.org.

RAND Feinberg Center for Catastrophic Risk Management and Compensation

The Feinberg Center, part of RAND Education, Employment, and Infrastructure, seeks to identify and promote laws, programs, and institutions that reduce the adverse social and economic effects of natural and manmade catastrophes by improving incentives to reduce future losses; providing just compensation to those suffering losses while appropriately allocating liability to responsible parties; helping affected individuals, businesses, and communities to recover quickly; and avoiding unnecessary legal, administrative, and other transaction costs.

For more information about the Feinberg Center, visit www.rand.org/feinberg-center or contact the director at ccrmc@rand.org.

Funding

Institute for Civil Justice work is supported by pooled grants from a variety of sources, including corporations, trade and professional associations, individuals, government agencies, and private foundations.

Feinberg Center work is supported by the generous contributions of the RAND Feinberg Center for Catastrophic Risk Management and Compensation Advisory Board, gifts from RAND supporters, and income from the operation of RAND Education, Employment, and Infrastructure.

Acknowledgments

We thank RAND's Institute of Civil Justice (ICJ) and RAND's Feinberg Center for Catastrophic Risk for their generous support. We also thank Oliver Brew, Bryan Choi, Philip Dawson, Emily Lowe, Michael von Gablenz, Erin Kenneally, Darren Pain, Ed Pocock, Karthik Ramakrishnan, Cristian Trout, Peter Wedge, Daniel Woods, ICJ board members, and other reviewers for their insightful comments and feedback.

We thank Lloyd Dixon and Benjamin Master for providing helpful feedback about the report and Babitha Balan and Phyllis Gilmore for overseeing the publication process. In addition to engaging expert human peer reviewers, the quality assurance team used assistive AI tools to suggest writing improvements to the authors.

We would also like to thank Connor McCarrick of Filing Focus and Robert Brauneis of George Washington University for access to insurance filing and lawsuit data.

The participation and inclusion of individuals in the acknowledgements does not imply their endorsement of the views, findings, or conclusions presented in this report.

Contents

About This Report iii
Figures and Tables vii

CHAPTER 1
Introduction 1

CHAPTER 2
Background 5
AI and Cyber: Similar Technology, Different Insurance Challenges 6
Research Methods 7

CHAPTER 3
AI Incidents, Lawsuits, and Laws Reveal Distinct Loss Mechanisms 9
Developers, Vendors, and Deployers Face Different AI Risks 9
Misinformation and Deepfakes Dominate Reported AI Incidents 10
AI Litigation Centers on Intellectual Property and Training Data 11
Most Enacted State AI Laws Target Harmful Synthetic Content 12
Six Root Causes Drive Most AI Losses 13

CHAPTER 4
How Insurers Are Responding to AI Risk 15
Major Carriers Are Excluding AI Losses 15
Some Carriers are Affirmatively Covering AI Losses 17
The Remaining Carriers Are Silent on AI Coverage 20

CHAPTER 5
Underwriting AI Risk Requires Combining Imperfect Tools 21

CHAPTER 6
Five Mechanisms That Could Produce AI Accumulation Risk for Insurers 23

CHAPTER 7
Conclusion and Recommendations 25
Conclusion 25
Recommendations 25
In Summary 28

APPENDIX
Insurance Lines 29

Abbreviations 31
References 33

Figures and Tables

Figures

2.1. Capability Evolution of AI 5
3.1. Multilabel Classification of Generative AI Incidents 11
3.2. Primary AI Lawsuit Focus Areas 12
3.3. AI Legislation 13
4.1. AI Perils and Insurance Lines 16
4.2. AI Exclusions in Admitted Policies 18
7.1. Relationship Between AI Cover, Insurer Accumulation Exposure, and Protection Gap 26

Tables

3.1. AI Harm Types 14
A.1. Descriptions of Insurance 29

CHAPTER 1

Introduction

Generative artificial intelligence (GenAI) is moving rapidly from an experimental tool to core business infrastructure. Firms are adopting machine learning, GenAI, and increasingly agentic systems to automate customer service, support coding and document review, make employment and lending decisions, generate marketing content, and control physical systems.

However, as of this writing in 2026, firms are only beginning to understand and appreciate the opportunities—and risks—of this emerging technology. Harms, losses, and lawsuits have already begun to materialize. For example, one user tricked a chatbot into selling a new sport utility vehicle for $1 (Tsuki, 2025). In another case, a company brought a civil claim against Google after its AI Overview web browsing feature incorrectly stated that the company was facing fraud charges (Martin, 2025). Further legal action includes a wrongful death and product liability lawsuit alleging that a defectively designed, emotionally manipulative chatbot contributed to the suicide of a 14-year-old boy (*Garcia v. Character Technologies, Inc.*, 2024).

Insurance offers an opportunity for firms to transfer these risks. But AI losses do not map cleanly onto existing insurance categories. Unlike familiar perils, AI-related losses can arise from faulty—but expected—model behavior, malicious misuse, rights violations, or governance failures. A single AI event may simultaneously resemble a cyber incident, a professional error, and a product defect. Unsurprisingly, insurers are diverging in their approaches to AI risk, with some carriers affirmatively covering AI losses across certain policy lines, others excluding it from their policies, and still others remaining silent. At the same time, specialized carriers and managing general agents are beginning to offer stand-alone AI coverage.

This report was motivated by the emerging mismatch between rapid enterprise adoption of AI and the fragmented market for insuring (or not insuring) AI use, particularly in the United States. Although public discussions of AI risk sometimes focus on extreme and speculative harms, the immediate challenge is more practical. Firms already face AI-related incidents and litigation, while insurers, brokers, regulators, and policyholders lack a shared understanding of which losses are occurring, how they should be classified, which lines should respond, and how correlated losses might accumulate across portfolios. The problem is compounded by sparse claims data, heterogeneous evaluation methods, growing use of exclusions, and the concentration of AI dependencies in cloud and hardware providers.

To address these issues, we examine AI insurability in four ways in this report. First, we analyze how AI-related incidents, lawsuits, and enacted state laws are emerging. Second, we examine how insurers are responding to these incidents through exclusions, endorsements, and affirmative cover. Third, we review current methods for assessing and underwriting AI risk. Fourth, we consider mechanisms by which AI could produce correlated or accumulated losses. Finally, we develop recommendations for insurers and reinsurers, brokers, and regulators for reducing the ambiguity of AI coverage to better track and manage the protection gap and accumulation exposure.

Our main findings and recommendations, summarized below, are relevant for insurers, reinsurers, brokers, and managing general agents, as well as for state insurance regulators and policymakers, including the National Association of Insurance Commissioners (NAIC).

Examining AI incidents, lawsuits, and enacted state laws, we found the following:

- Of public GenAI incidents, 84 percent (n = 599) relate to *misinformation* (false, deceptive, or manipulative content) and *deepfakes* (AI-generated or altered media impersonating a real person or event).
- Of U.S. litigation for GenAI harms, 60 percent (150 of 249) relate to the alleged violations of intellectual property (IP) or improper training by AI developers rather than to the use, misuse, or outputs of these models.
- Most U.S. states have adopted laws focusing on restricting harmful synthetic content, such as nonconsensual intimate images (52 state laws) and child sexual abuse material (CSAM) (30 state laws).

Examining how insurers are responding to AI risks, we found the following:

- Unlike cyber, AI risks are not contained to a single insurance line but span many lines: technology errors and omissions (E&O), professional liability, cyber, directors and officers, commercial property, and more. Moreover, proper management of AI-related losses will require approaches that may draw on cyber experience but will extend far beyond it.
- Insurance carriers are responding to AI risks unevenly, with some offering affirmative coverage for certain risks (e.g., model errors), and others filing broad exclusions for any AI-related loss.
- Most carriers are remaining silent, neither affirmatively covering nor excluding AI-related risks in their policies, leaving coverage for such losses untested and open to dispute.

We identified five mechanisms by which AI could expose insurers to large, concentrated financial losses:

- *Universal AI attacks* are adversarial attacks that exploit the same vulnerability across many AI systems. This is analogous to cyber risk in which the same hardware or software vulnerability exposes many companies.
- *Common model dependency* occurs when many companies rely on the same hardware, software, or infrastructure component.
- As *force multipliers*, AI tools can augment cyberattacks at a greater scope and scale than previously possible.
- A *legal shock* is a legal ruling or action that suddenly holds many companies liable for past or current actions or triggers regulatory enforcement against noncompliant firms, resulting in many insurable losses.
- *Subtle corruption or degradation* describes the possibility of AI models suffering subtle or prolonged corruption or degradation of model performance or behavior that leads to more-complicated but correlated failures.

The most important near-term policy challenge is not whether to insure all AI risks but whether policyholders and carriers share a common understanding of what is covered, what is excluded, and where correlated exposures may be growing. We therefore focus on three practical priorities for reducing the ambiguity around AI coverage to better manage the national protection gap and the accumulation exposure within insurance markets:

- **Clarify AI coverage.** State insurance regulators and the NAIC, supported by carriers and brokers, should develop an AI Coverage Notice enabling carriers to clarify how, and under which lines, AI-related losses are affirmatively covered, excluded, or left silent.

- **Standardize incident reporting for attritional and accumulation loss.** Policy researchers and brokers, together with carriers and reinsurers, should develop a common, adaptable taxonomy for AI incidents, claims, and controls to properly track and report on AI risks. Actual incident data should remain private, however.
- **Manage accumulation risk.** Insurers and reinsurers should analyze AI aggregation scenarios, focusing on shared model dependency, common infrastructure failure, and regulatory shock to manage correlated AI risk.

The remainder of this report proceeds as follows. Chapter 2 describes our background, data sources, and research methods. Chapter 3 examines how AI incidents, lawsuits, and enacted state laws reveal distinct loss mechanisms and identifies the root causes driving most AI losses. Chapter 4 turns to how insurers are responding—through exclusions, affirmative coverage products, and silence. Chapter 5 reviews the imperfect tools available for underwriting AI risk. Chapter 6 considers five mechanisms that could produce accumulation risk for insurers. Chapter 7 concludes with our recommendations.

CHAPTER 2

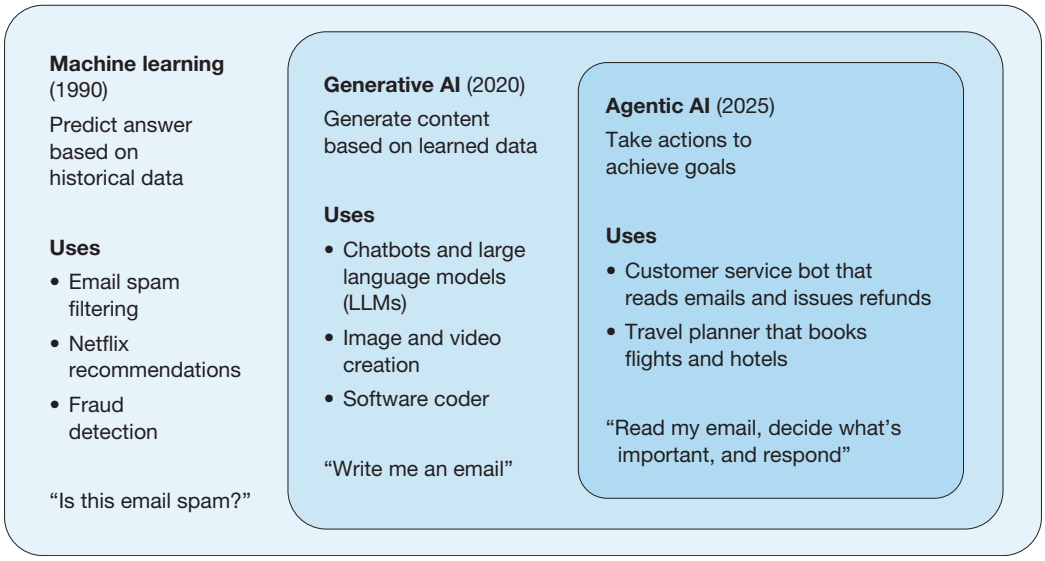
Background

AI has evolved greatly over the past 20 years, and especially in the past ten years, as shown in Figure 2.1.

Machine learning is a group of algorithms that learn patterns from data. Machine learning adoption expanded through two distinct phases, progressing from early statistical pattern identification to advanced deep-learning networks (Mitchell, 1997). Deployment began in the late 1990s and was driven by internet connectivity, increased computational power, and the formalization of statistical models, including support-vector machines and random forests (Jordan and Mitchell, 2015). Deep learning and neural networks initiated a second period of accelerated growth in the mid-2010s that showed vastly improved performance and capabilities over previous machine learning techniques (LeCun, Bengio, and Hinton, 2015).

GenAI is "a machine-based learning system or model that is trained on data with the ability to create content or responses, including but not limited to text, images, audio, video, or code" (LaMorte, 2026).[1] Early probabilistic classifiers and deep-learning frameworks, including Naive Bayes, variational autoencoders, and generative adversarial networks, introduced early forms of data generation (for a helpful overview, see Stryker, undated). However, feature assumptions and limited context constrained these approaches. Transformer architectures revolutionized GenAI, such as Generative Pre-trained Transformer (GPT). The limited release of GPT-3 in June 2020 marked what many consider the start of the modern GenAI era—although it was ChatGPT's launch in November 2022 that brought the technology to mass adoption.

FIGURE 2.1
Capability Evolution of AI



[1] This is the definition that the Insurance Standard Organization uses, although we recognize that other carriers may use different but qualitatively similar definitions.

Unlike GenAI systems, which primarily respond to prompts with content, agentic systems are designed to pursue goals over multiple steps, make intermediate decisions, and interact with external tools or environments with limited human intervention (Acharya, Kuppan, and Divya, 2025). This added autonomy expands their usefulness for complex business processes but also increases the potential for operational errors and unintended actions and raises more-difficult questions about accountability and liability.

AI and Cyber: Similar Technology, Different Insurance Challenges

Cyber insurance may seem like a natural reference point for managing AI-related losses because both involve technology-driven, intangible, and rapidly evolving risks. However, AI risk differs from cyber risk in several important ways that shape how insurers underwrite AI risk, how they allocate losses across insurance lines, and how coverage disputes are likely to emerge.

First, **cyber underwriting** focuses on technical controls, such as firewalls, patch management, and incident response (Kenneally, 2021). By contrast, as we discuss below, *AI risk underwriting* focuses more on governance, including model validation, data management, and human oversight ("human in the loop"). This implies a shift from assessing perimeter defenses and security hygiene to assessing organizational controls and model assurance practices.

The **nature of the underlying risk** also differs. Human actors, whether malicious outsiders or negligent insiders, typically cause cyber losses, often acting through an identifiable trigger, such as a data breach or ransomware attack (Pain and Romanosky, 2026). By contrast, AI losses may arise from expected—but unintended—system operations, such as hallucinations (plausible but factually incorrect), from biased decisions, or from an AI agent acting unexpectedly to cause damage. As a result, causation is often less straightforward in AI contexts because the source of loss may lie in the model itself, the training data, the deployment environment, the surrounding controls, or the user's reliance on the output. This makes AI losses harder to attribute than cyber events and complicates claims adjustment and liability allocation.

> AI risk differs from cyber risk in their approaches to underwriting and regulation, the nature of the underlying risk, and social and ethical considerations.

In addition, the bulk of cyber insurance claims relate to first-party losses (e.g., losses to the firm from fraud, the cost of investigation, the repair of IT systems, etc.) (see Coalition, 2026, Figure 1.3), while most AI losses to date relate to third-party claims from litigation.

The **regulatory environment** also differs between cyber and AI. Regulation for cyber risk (such as mandated requirements and the enforcement actions that state or federal governments have taken) has evolved over the past decades to center on privacy, consumer protection, and data-breach reporting (Hartzog and Solove, 2022). AI regulation, on the other hand, focuses on criminalizing generation of harmful content and on regulating biased output and automated decisionmaking through transparency requirements. For insurers and policyholders, this represents a material difference in compliance obligations.

Finally, AI introduces persistent **ethical and social dimensions** that are less applicable in cyber insurance. Cyber losses are primarily economic and privacy-related, but AI-related losses can involve discrimination, misinformation, workforce displacement, psychosocial dislocation and behaviors (e.g., addiction), harmful parasocial relationships, and erosion of public trust.

Taken together, these differences suggest that AI insurance will require approaches that may draw on cyber experience but will extend far beyond it.

Research Methods

Literature Review

We conducted a targeted literature review to identify published work on AI risk and AI insurance. The review covered both industry and academic sources and was designed to capture how insurers, reinsurers, brokers, modeling firms, and policy researchers are currently framing AI-related losses, coverage responses, underwriting practices, and accumulation concerns. We sought reports, white papers, market briefings, and research papers that discussed AI-related harms, insurance coverage, exclusions, underwriting, model evaluation, or accumulation risk. We also reviewed publications cited within these documents to identify additional relevant sources. Because the market is changing quickly, we prioritized recent materials and focused mainly on publications from 2023 through mid-2026, which corresponds with the period following the broad adoption of GenAI tools after the launch of ChatGPT.

We screened sources for relevance based on whether they addressed one or more of the following: (1) AI harms or loss mechanisms, (2) insurance coverage or exclusions for AI-related losses, (3) underwriting or evaluating AI risk, (4) accumulation or systemic risk arising from AI, or (5) legal or regulatory developments affecting AI insurability. We excluded publications that focused solely on technical AI performance without implications for risk transfer, liability, or insurance and purely promotional material lacking substantive discussion of insurance or risk. This process resulted in a final review set of 41 papers published between 2023 and 2026.

Data Analysis for AI Incidents, Lawsuits, State Laws, and Insurance Filings

To our knowledge, the most comprehensive and publicly available data related to AI harms come from two sources: a crowdsourced dataset of actual AI incidents and George Washington University's dataset of AI-related lawsuits.[2]

Our information on *AI incidents* (alleged harm or near harm to people, property, institutions, or the environment facilitated by an AI system) came from the AI Incident Database (AIIDB) (McGregor, 2021). Launched in 2020, the AIIDB is a volunteer community effort seeking to document "failings of intelligent systems" and make them "discoverable by developers, researchers, and policy makers" (Responsible AI Collaborative, 2022).[3] As of June 20, 2026, this database contained more than 1,537 incidents, dating back to 1988.[4] In this report, we consider the 713 GenAI incidents that occurred after the release of ChatGPT on November 30, 2022. The AIIDB represents the most comprehensive incident data currently available; however, rigorous validation has not yet been completed. We therefore treated this database as an indicator of the types and frequency of AI harms rather than an authoritative record.

Our information on AI lawsuits came from the George Washington University Database of AI Litigation (DAIL) (Ethical Tech Initiative, undated). DAIL uses broad inclusion criteria to collect AI lawsuits related to algorithmic hiring and credit, GenAI training, and authorship of AI. This database does not include AI

[2] Testudo, an insurance provider, also has curated a set of AI lawsuits; however, these data are not public.

[3] We acknowledge that other datasets exist, such as AIAAIC, 2026; Organisation for Economic Co-operation and Development Artificial Intelligence, 2026; MIT AI Risk Initiative, undated; and AI Vulnerability Database, undated. However, AIIDB is most relevant to our context.

[4] The AIIDB was inspired by analogous databases in aviation safety and cybersecurity; its mission is to "identify, define and catalogue artificial intelligence incidents" (Responsible AI Collaborative, 2022, p. 5). It defines *incidents* as cases in which intelligent systems caused safety, fairness, or other real-world harms. The database is open source, governed by the Responsible AI Collaborative, and receives funding from private sector companies and philanthropic organizations (Responsible AI Collaborative, 2022).

patent litigation or securities class actions involving misrepresentations by firm executives about AI use. As of June 16, 2026, DAIL contained 457 lawsuits filed between 2004 and 2026. The collection primarily includes federal lawsuits ($n = 330$) because of the federal jurisdiction over copyright and IP. State-level lawsuits ($n = 68$) are also included. These capture common law torts and primarily involve AI companions, that is, conversational AI systems designed to simulate personal, emotional, or relational interaction with users. The international lawsuits are limited ($n = 59$) by translation challenges, restricted access to foreign dockets, and a lack of online transparency laws in foreign jurisdictions. While other AI litigation data sources exist, DAIL is the only one of which we are aware that allows full access for analysis. Here, we have constrained the analysis to the 249 lawsuits classified as GenAI in the dataset.

In addition, we examined U.S. state legislation related to civil and criminal penalties for AI-related harms. We used information from the U.S. AI Law Tracker–All States (Orrick, undated-b). This law tracker collects information on U.S. state criminal legislation and civil statutes that have been enacted (i.e., not proposed laws). As of June 20, 2026, this dataset contained 234 individual state laws. We then manually excluded laws without any distinct penalty (i.e., laws related to compliance, educational directives, or establishing oversight boards). This left us with 189 state laws enacted between 2006 and June 2026

Finally, we examined regulatory insurance forms in the U.S. admitted market related to AI. We retrieved endorsements and exclusion forms from FilingFocus, a commercial platform that provides structured, searchable access to the System for Electronic Rate and Form Filing (SERFF), the electronic platform insurers use to submit rate and form filings to state regulators. We obtained forms using keyword queries for AI, including *llm*, *machine learning*, and *generative ai*. We combined near-duplicate forms (those with identical group, action category, filing count, submission dates, and definition text but distinct file URLs), reducing the number of forms in the analysis to 125, spanning from the release of ChatGPT through June 21, 2026. We determined insurance lines for each filing by cross-referencing SERFF Sub–Type of Insurance numeric codes and keyword matches against form titles extracted from the PDF header. For example, "Commercial Umbrella & Excess" was mapped to SERFF code 17.1020 and identified through such terms as *commercial umbrella*, *excess liability*, and *specialty umbrella*. When neither source produced an assignment, the insurance line is labeled *Other*. The final dataset includes 112 exclusion forms, four applications, and seven endorsements, submitted by 60 distinct insurance companies.

Expert Interviews

We conducted 22 interviews with 26 insurance and legal scholars between February and March 2026. In selecting interviewees, we used a nonprobabilistic convenience-sampling strategy (Guest, Bunce, and Johnson, 2006), identifying and recruiting individuals through LinkedIn, the research team's professional networks, and targeted online searches of relevant organizations. We selected individuals for outreach based on their subject-matter expertise in AI, insurance, or technology. We also employed the chain (snowball) sampling approach (Goodman, 1961), inviting interviewees to recommend other potential participants.

Participation was voluntary, and we provided no financial incentives. The research protocol received approval from our institutional review board. We conducted interviews both in person and remotely via video conferencing, and these generally lasted approximately one hour. We took notes and did not record the interviews. We used a semistructured interview guide and analyzed data using thematic analysis to identify and document key challenges and considerations associated with insurance of AI technologies (Braun and Clarke, 2006; Glaser and Strauss, 2017).

In the course of this research, we used RAND's instances of ChatGPT 5.4 and Anthropic's Claude code to revise small segments of written content, analyze and classify data, validate and proofread research sections, and write python programs to visualize data. The authors reviewed any AI-generated content.

CHAPTER 3

AI Incidents, Lawsuits, and Laws Reveal Distinct Loss Mechanisms

Businesses use AI tools in many applications across all industry sectors. However, because of the unpredictable (stochastic) behavior of generative models, they can produce incorrect outputs or cause damage, leading to financial, physical, and property harms. For example, consider the following sample of AI use cases, each producing different forms of harm and loss:

- an AI financial lending model or employee screening tool producing discriminatory results, leading to third-party claims against the firm
- cyber criminals using AI to find and exploit new software vulnerabilities, leading to business interruption for a firm that was otherwise well protected
- an autonomous warehouse vehicle receiving a faulty AI software update, causing bodily injury to employees and/or property damage inside the warehouse
- an AI agent deleting a firm's production software systems, causing business interruption[1]
- AI being used to generate marketing material that produces images that infringe on another firm's copyrighted works
- law enforcement using an AI image recognition tool used that misidentifies a suspect, leading to third-party claims
- an AI decision tool providing incorrect market analysis or financial figures, causing share prices to drop and resulting in investors filing a securities class-action lawsuit.

Developers, Vendors, and Deployers Face Different AI Risks

The risks, exposures, and potential liabilities associated with AI vary depending on a company's role within the AI supply chain. Stakeholders generally fall into three categories: developers, vendors, and deployers.

AI *developers* include frontier labs pushing the state of the art to create and train foundation (general-purpose) AI models, such as Anthropic (Claude), OpenAI (GPT), and Google (Gemini); open-source, open-weight model developers, such as Meta (Llama), Mistral, and DeepSeek; and smaller or specialized model developers.[2] Developers primarily face risks from IP disputes and privacy concerns arising from scraping and using training data. Developers also face potential liability for negligence if a failure to design, test, or implement systems with reasonable care leads to harm or technical errors, such as hallucinations. Because they rarely control how models operate in specific real-world contexts and typically attempt to limit liability

[1] An *AI agent* is an autonomous software system (program) that uses AI to perceive its environment, make decisions, and execute multistep actions to achieve a specific goal.

[2] *Open-source AI models* provide the complete recipe for building the model, including code and training data; *open-weight models* provide only the final numerical parameters used to generate output (the weights).

through contracts, developers' financial exposure remains narrow and highly episodic. However, whether the courts will support these limits in the event of large-scale harms is an open question.

Vendors package, integrate, or embed AI models into commercial products and services, including such products as Adobe Firefly and Notion AI. The legal and financial risks for vendors center on negligence, the misrepresentation of product capabilities or performance, and contractual breaches.

Deployers leverage AI models and products for specific use cases ranging from customer service and fraud detection to insurance underwriting and candidate assessment. These enterprise deployers may be the ones most exposed to operational, regulatory, and end-user harms resulting from implementation decisions.

These use cases illustrate the kinds of losses AI may generate and, therefore, the risk profiles. To better understand the insurable risk landscape across these stakeholders, we next examine three empirical sources of evidence: publicly reported AI incidents, filed lawsuits, and enacted state laws. Incidents provide a near-term view of observed failures; lawsuits identify harms that have matured into legal claims across these stakeholder groups; and legislation highlights the types of harms and losses that regulators and lawmakers increasingly treat as actionable.

Misinformation and Deepfakes Dominate Reported AI Incidents

As noted earlier, the best available data on harms caused by AI systems come from the AIIDB, which included 713 GenAI incidents at the time of this writing. Figure 3.1 illustrates the types of harms using a multilabel classification that assigns each incident to one or more categories.[3]

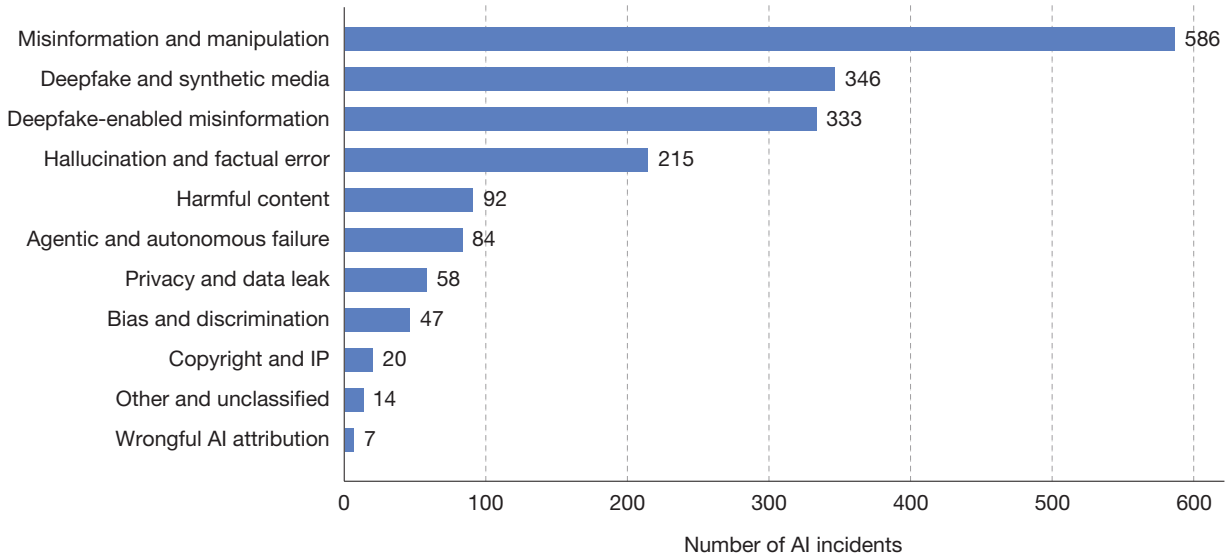
Given these data, misinformation and deepfakes accounted for 84 percent (n = 599) of GenAI incident reports in AIIDB. Misinformation incidents describe situations in which AI generates or amplifies false, deceptive, or manipulative content, such as scammers using AI to generate responses to California community college applications to fraudulently collect federal financial aid (Lurye, 2025). Deepfakes relate to AI-generated media that impersonate, harass, or deceive; examples include AI-generated images of Pope Francis in a white puffer jacket (Huang, 2023) and sexually explicit images of Taylor Swift ("Deepfake Images of Taylor Swift Spread," 2024). Over 50 percent of incidents categorized as misinformation and deepfakes intersect, indicating that synthetic media (i.e., fully AI-generated content) is frequently the mechanism of manipulation; one example is deepfake audio of a leading Slovakian politician discussing how to rig the vote, which circulated two days before Slovakia's 2023 parliamentary election (Zuidijk, 2023).

> The most common types of AI incidents are those involving false, deceptive, or manipulated content.

Additionally, 47 percent of incidents fall into the categories of (1) hallucination, (2) safety and harmful content, or (3) agentic and autonomous failure categories. Here, we categorized incidents involving AI systems producing confidently stated but materially false outputs as hallucinations and factual errors (n = 215, 30 percent); one example is a Binnall Law Group attorney who filed a motion containing phantom legal cita-

[3] To classify AI incidents, we considered both keywords and LLM classifications. Keyword classification ensures that classifications remain consistent and auditable as incidents are added to the database. However, this system is brittle and less generalizable to the choice of keywords, which are derived from the language of past incidents. In contrast, the LLM approach captures semantic context that keyword matching misses but is sensitive to the design of the classification taxonomy. However, classifications may shift across runs or with model updates, raising replicability concerns and reducing interpretability. Here, we have leveraged keyword search to prioritize reproducibility and auditability as the database expands. The keywords for each harm category were based on common terms found in the incident summaries; for example, the hallucination classification includes *hallucination*, *fabrication*, *false information*, *misinformation*, and *fictitious*, while the deepfake and synthetic information classification includes *deepfake*, *fake image*, *fake video*, and *voice cloning*.

FIGURE 3.1
Multilabel Classification of Generative AI Incidents



SOURCE: Features information from the AI Incident Database.
NOTE: These data reflect the types of harms caused by 713 unique incidents, although an incident may be classified into multiple categories.

tions generated by Claude (Purifoy, 2026). Incidents categorized as related to safety and harmful content (n = 92, 13 percent) cover dangerous or psychologically harmful AI outputs, including the situation in which the National Eating Disorders Association shut down its chatbot after it gave weight-loss advice to users seeking support for eating disorders (Wells, 2023). Agentic and autonomous failures (n = 84, 12 percent) include incidents of AI systems taking unintended, unauthorized, or irreversible real-world action. For example, a Cursor AI agent deleted PocketOS's production database and backups (Carter, 2026).

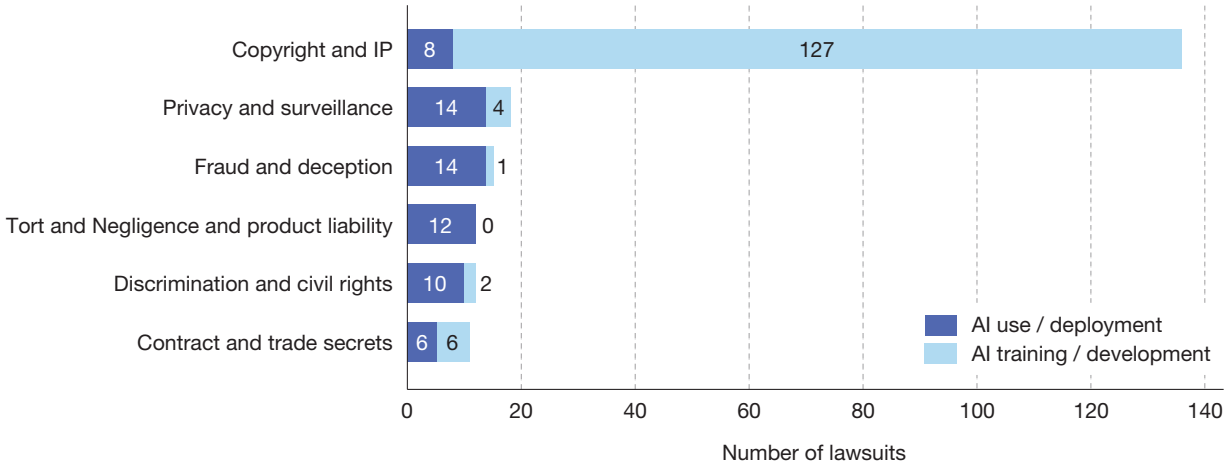
Additionally, we analyzed the form of GenAI used in each incident, including the text, audio, video, and code generation. Audio and video generation appear in the largest share of incidents, accounting for 32 percent (n = 226) and 25 percent (n = 181) of the incidents, respectively.

AI Litigation Centers on Intellectual Property and Training Data

Next, we examine the primary focus of U.S. based lawsuits to date relating to GenAI, as collected by the Ethical Tech Initiative at George Washington University (Figure 3.2).[4] These data suggest that the vast majority of GenAI litigation relates to the training (creation) of AI models by AI developers rather than the use or misuse of outputs of these models. Copyright and IP claims regarding training data ingestion account for nearly two-thirds of the suits (150 of 249 lawsuits). Plaintiffs predominantly allege the unauthorized and uncompensated use of protected content to build foundation models. For example, *The New York Times* has alleged that OpenAI and Microsoft used millions of its copyrighted articles to train ChatGPT and Copilot without permission or payment (*The New York Times Company v. Microsoft Corporation et al.*, 2023).

[4] We based the keywords for each category on common terms in the case summaries and related laws. For example, the copyright and IP classification includes *trademark*, *patent*, *infringement*, *17 U.S.C. 1202*, and *Digital Millennium Copyright Act*, while the privacy and surveillance classification includes *Biometric Information Privacy Act*, *BIPA*, *Health Insurance Portability and Accountability Act*, *HIPAA*, *18 U.S.C. 2510*, and *18 U.S.C. 1030*.

FIGURE 3.2
Primary AI Lawsuit Focus Areas



SOURCE: Features information from the DAIL by the Ethical Tech Initiative at George Washington University.

Enterprise use of GenAI is also driving deployment-related litigation. Privacy and surveillance cases (n = 18) center on unauthorized data collection and interception, including covert third-party tracking embedded in AI platforms, unauthorized scanning of user content, biometric surveillance, and wiretapping violations. For example, in *Thele v. Google LLC (*2025), the plaintiff, Thomas Thele, alleges that Google activated its Gemini AI to scan user content, in Gmail, Chat, and Meet, without consent (Roy, 2025).

Fraud and deception cases (n = 15) reflect two distinct patterns. In the first, a company engages in "AI washing" by making materially false or misleading representations about their AI capabilities to investors and consumers. In *Dolly v. GitLab Inc.* (2025), shareholders alleged that executives overstated the performance of the company's AI coding assistant to inflate its market position. The second pattern involves AI-enabled harms, including financial scams, impersonation, and unauthorized exploitation of creative work. For example, in *United States v. Smith* (2024), the defendant allegedly used AI-generated music to fraudulently collect more than $10 million in streaming royalties.

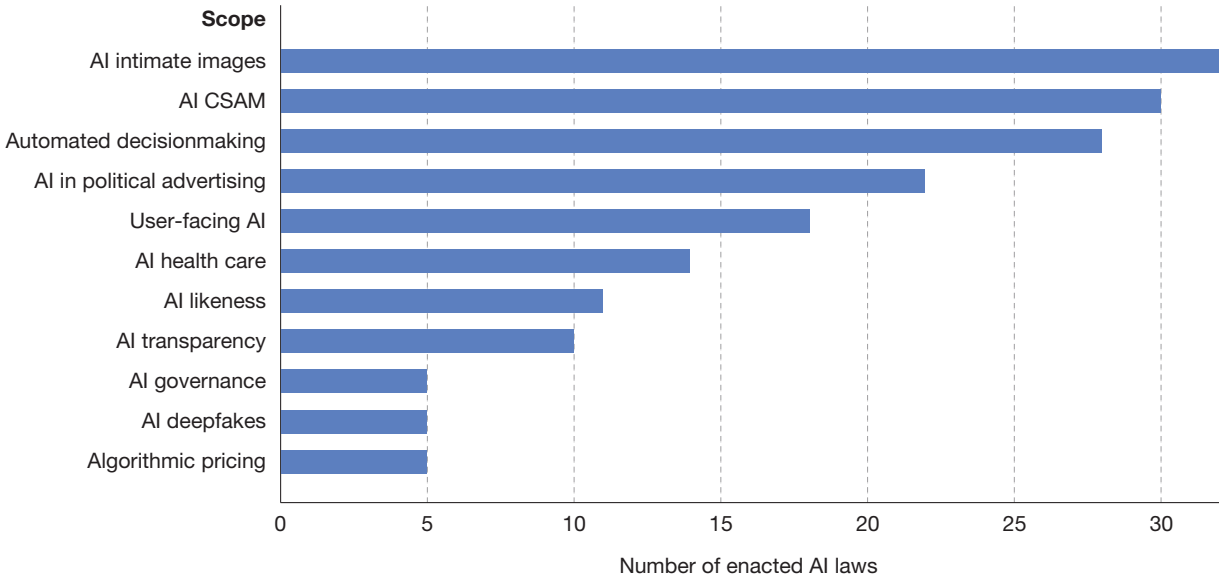
Tort and product liability cases (n = 12) center on inadequate safeguards and manipulative design that result in product defects producing false or harmful recommendations or harmful material. For example, in *Garcia v. Character Technologies* (2024) and *Raine et al. v. OpenAI, Inc.* (2025), plaintiffs alleged that a software design defect in companion or frontier AI models contributed to the suicide or suicidal ideation of minors. In other cases, including *Doe v. X.AI Corp.* (2025), plaintiffs alleged strict product liability for AI systems that generate nonconsensual sexually explicit imagery.

Most Enacted State AI Laws Target Harmful Synthetic Content

Next, we examine civil and criminal legislation enacted across the U.S. states, as shown in Figure 3.3. Most enacted U.S. state AI laws focus on restricting harmful synthetic content. Legislatures across 33 states have passed a total of 62 laws targeting nonconsensual intimate images and/or addressing CSAM. These statutes expanded existing criminal codes to incorporate synthetic depictions and criminalize the production and distribution of materials that are indistinguishable from, or realistically depict, a

The majority of state laws address harmful synthetic content concerning nonconsensual intimate images and CSAM.

FIGURE 3.3
AI Legislation



SOURCE: Features information from the Orrick, undated-a. Please see webpage for most current information.

minor or child. Regulatory frameworks also address civic and personal depiction risks through 22 political advertising laws, 11 likeness laws, and five deepfake laws. These statutes mandate disclaimers for AI-altered campaign content, ban deceptive election deepfakes, establish civil enforcement mechanisms, and enforce consent-based prohibitions against unauthorized digital replicas.

State regulatory frameworks increasingly govern automated and AI-driven decisionmaking systems to protect consumer rights and professional boundaries. A core group of 28 automated decisionmaking laws establishes consumer opt-out rights, mandates data protection impact assessments, requires employment bias audits, and limits sole reliance on AI in health care prior-authorization processes. Consumer protections extend to interactive contexts through 18 user-facing AI laws that require chatbots to disclose their AI identity, implement self-harm detection protocols, and prohibit unlicensed AI mental health services. Health care operations face additional constraints from 14 specific laws that prohibit automated systems from independently issuing adverse coverage determinations or simulating licensed clinical credentials. Finally, 15 combined transparency and governance laws enforce disclosure labels, require provenance metadata, mandate risk management programs, and apply antitrust restrictions to algorithmic pricing tools to prevent coordinated pricing.

Six Root Causes Drive Most AI Losses

Taken together, these publicly reported incidents, lawsuits, and legislative developments identify the mechanisms (root causes) leading to AI losses and harms. Identifying these mechanisms matters for determining whether and how AI losses fit within insurance lines. First, we observe that many incidents are due to model error, such as when the model produces unwanted, faulty, or factually incorrect output. These incidents are the result of expected—but unwanted—model behavior.[5]

[5] For example, AI experts concede that model errors, such as hallucinations, are unwanted but are an expected behavior of the AI (transformer) operation.

Next, AI losses can also arise from security vulnerabilities when a system's guardrails are missing, weak, or bypassed. Unlike malicious use of a functioning model, the harm here stems from the system's own inadequate safeguards, for example, a jailbreak or prompt-injection attack that defeats safety controls or a backdoored model distributed through a public repository.

As the lawsuits suggested, many incidents are due to violations of informational (i.e., copyright, IP), personal, or civil rights that stem from deliberate ways in which an AI model was trained (e.g., on personal or copyrighted material) or how it is applied (e.g., hiring, financial decisionmaking, or generating content that infringes on copyrighted material).

GenAI losses can also occur, of course, from malicious use of the model to cause fraud or deception, such as when it is used to produce deepfake images or videos or to create a phishing email leading to corporate fraud. Employees of deployers or third parties can also maliciously use models hack into or gain control of the AI model and deploy it for improper or illegal purposes.

Finally, bad governance can cause AI incidents resulting from mismanagement of AI tools within an organization or improper communication from executives about these tools to stakeholders. Table 3.1 summarizes these mechanisms.

Given these AI harm types and their characteristics, we next examine how insurers respond to cover AI losses.

TABLE 3.1
AI Harm Types

Loss or Harm Type	Harm Mechanism	Description of Harm
Model error	Expected but unwanted model output	AI produces incorrect, misleading, or deficient outputs, for example, a customer-service chatbot providing incorrect advice or an autonomous robot causing an accident or injury
Security vulnerability	Inadequate safeguards	Inadequate guardrails let an adversary exploit the AI, for example, a jailbreak defeating safety controls or a poisoned model being downloaded from a public model hub
Information or copyright violation	AI training or use	AI infringes on protected or creative works, for example, an AI model producing copyrighted output or being trained on copyrighted material
Civil or personal rights violation	AI use	AI infringes on protected personal rights, for example, an AI hiring model producing discriminatory outcomes
Fraud or deception	Malicious use or intent	AI is used to deceive, impersonate, or commit criminal acts, for example, creating deepfakes or powering phishing attacks
Bad governance	Management misconduct	Losses arise from misleading disclosure, or misrepresentation about AI capabilities or controls, for example, a company exaggerates its AI capabilities and later faces securities litigation or regulatory scrutiny

CHAPTER 4

How Insurers Are Responding to AI Risk

The many forms of loss that GenAI can produce raise a fundamental question for insurers: which lines of insurance, if any, should respond. Indeed, many insurance stakeholders have presented their views on the mapping between AI perils and insurance lines (Munich RE, 2024; Jia, 2025; AON, 2025; Kenneally, 2021; Lloyd's, 2025; Hill and Pocock 2026; Brew and Fall-Bailey, 2026). For example, Figure 4.1 identifies more than a dozen AI perils and 11 potential insurance lines that may respond. (Appendix A briefly describes these insurance lines.)

Figure 4.1 illustrates both the range of AI-related perils and the multiple insurance lines that may respond to a loss. Unlike more traditional perils, such as flood, fire, or even cyber, AI does not yet correspond to either a single peril or a stand-alone insurance line because the same model can generate very different forms and severity of insured loss depending on how it is trained, deployed, and used. As the incidents, lawsuits, and laws discussed in the previous chapter show, AI harms vary in mechanism, severity, latency (between incident occurrence, resulting harm, and filed claim), and legal posture. These characteristics help explain why AI does not map neatly onto a single insurance line and why insurers are responding through a mix of exclusions, endorsements, and affirmative products. Next, we examine these exclusions.

Major Carriers Are Excluding AI Losses

Unsurprisingly, some carriers are affirmatively excluding AI losses from many lines of insurance. In January 2026, Verisk/ISO, whose standardized policy forms appear in more than 80 percent of U.S. property and casualty policies (Polales Horton & Leonardi LLP, 2026), developed exclusionary language that carriers can adopt to exclude bodily injury, property damage, or personal and advertising injury caused by GenAI (CG 40 47, CG 40 48, CG 35 08) (Germond, 2025).[1] These commercial general liability exclusions are significant because they are specific to GenAI, and U.S. carriers use them broadly.

In addition, the insurance carrier Berkley revised its specialty lines of liability coverage, including directors and officers, E&O, and fiduciary liability policies, to exclude coverage with respect to "any actual or alleged use, deployment, or development of Artificial Intelligence" (see box on p. 19). This exclusion is quite broad in scope, written to bar coverage for claims connected to nearly any aspect of AI. It applies to claims arising from a company's use, development, or deployment of AI and from AI-generated content or communications. The exclusion also reaches losses tied to failures to identify AI-created material produced by others; inadequate or absent AI policies and training; and harms linked to AI-enabled products, services, chatbots, or virtual assistants. In addition, it excludes claims based on the company's own statements about AI, including how AI is used, how AI-related risks are managed, and what future AI plans or capabilities the company may have. It further bars coverage for claims involving alleged violations of AI-related laws or regulations, along with the costs of responding to legal or regulatory demands to assess, investigate, monitor, or address AI risks.

[1] However, note that we have no current information on the number of carriers, across states, that have actually adopted this language.

FIGURE 4.1
AI Perils and Insurance Lines

Peril	Media liability	Tech E and O, MPL, PI	Product liability	General liability	Intellectual property	Cyber	Crime	D and O	Employment
Faulty product or service	●	●	●	●	●	●	●	●	●
Copyright or trademark	●	●	●	●	●	●	●	●	●
Patent infringement	●	●	●	●	●	●	●	●	●
Discrimination and bias	●	●	●	●	●	●	●	●	●
Defamation, libel, slander	●	●	●	●	●	●	●	●	●
Bodily injury	●	●	●	●	●	●	●	●	●
Tangible property damage	●	●	●	●	●	●	●	●	●
Privacy and security breaches	●	●	●	●	●	●	●	●	●
Loss of financial assets/market manipulation	●	●	●	●	●	●	●	●	●
Product recall	●	●	●	●	●	●	●	●	●
Business interruption	●	●	●	●	●	●	●	●	●
Breach of D and O duties	●	●	●	●	●	●	●	●	●
Robotics	●	●	●	●	●	●	●	●	●
"Deepfakes"	●	●	●	●	●	●	●	●	●

● Available ● Limited ● Excluded unless custom contingent liability added

SOURCE: Reproduced from AON, 2025 (reproduced with permission).

NOTE: Bodily injury and tangible property damage are two perils without identifiable cover because general liability insurance, for example covers products, not professional services. MPL = miscellaneous professional liability; PI = professional indemnity; D&O = directors and officers liability.

There are additional reports that other major U.S. carriers, such as AIG, Great American, and Chubb, also exclude or are preparing to exclude AI losses (Gross, 2026; Intelligence Council, 2026a).[2]

The SERFF filing data reveal the introduction of AI exclusions in the admitted market (Figure 4.2). The data show a surge of exclusion activity beginning in the summer of 2025, understandably in response to the capabilities and risks of GenAI systems. Most exclusions relate to commercial umbrella or commercial general liability policies, suggesting broad carrier interest in excluding these losses from third-party lines with broad cover and high limits. In addition, carriers insuring construction companies are excluding third-party

[2] Interestingly, Chubb is seeking to exclude AI coverage at the same time that it has commented on how AI tools are helping improve insurance underwriting and claims processing (Intelligence Council, 2026b).

Berkley AI Exclusion Language

The **Insurer** shall not be liable to make payment under this Coverage Part for Loss on account of any Claim made against any **Insured** based upon, arising out of, or attributable to:

(1) any actual or alleged use, deployment, or development of **Artificial Intelligence** by any person or entity, including but not limited to:
   (a) the generation, creation, or dissemination of any content or communications using **Artificial Intelligence;**
   (b) any Insured's actual or alleged failure to identify or detect content or communications created through a third party's use of **Artificial Intelligence**;
   (c) any **Insured's** inadequate or deficient policies, practices, procedures, or training relating to **Artificial Intelligence** or failure to develop or implement any such policies, practices, procedures, or training;
   (d) any Insured's actual or alleged breach of any duty or legal obligation with respect to the creation, use, development, deployment, detection, identification, or containment of **Artificial Intelligence**;
   (e) any product or service sold, distributed, performed, or utilized by an **Insured** incorporating **Artificial Intelligence**; or
   (f) any alleged representations, warranties, promises, or agreements actually or allegedly made by a chatbot or virtual customer service agent;

(2) any **Insured's** actual or alleged statements, disclosures, or representations concerning or relating to **Artificial Intelligence**, including but not limited to:
   (a) the use, deployment, development, or integration of **Artificial Intelligence** in the **Company's** business or operations;
   (b) any assessment or evaluation of threats, risks, or vulnerabilities to the Company's business or operations arising from **Artificial Intelligence**, whether from customers, suppliers, competitors, regulators, or any other source; or
   (c) the **Company's** current or anticipated business plans, capabilities, or opportunities involving **Artificial Intelligence**;

(3) any actual or alleged violation of any federal, state, provincial, local, foreign, or international law, statute, regulation, or rule regulating the use or development of **Artificial Intelligence** or disclosures relating to **Artificial Intelligence**; or

(4) any demand, request, or order by any person or entity or any statutory or regulatory requirement that the **Company** investigate, study, assess, monitor, address, contain, or respond to the risks, effects, or impacts of **Artificial Intelligenc**e.

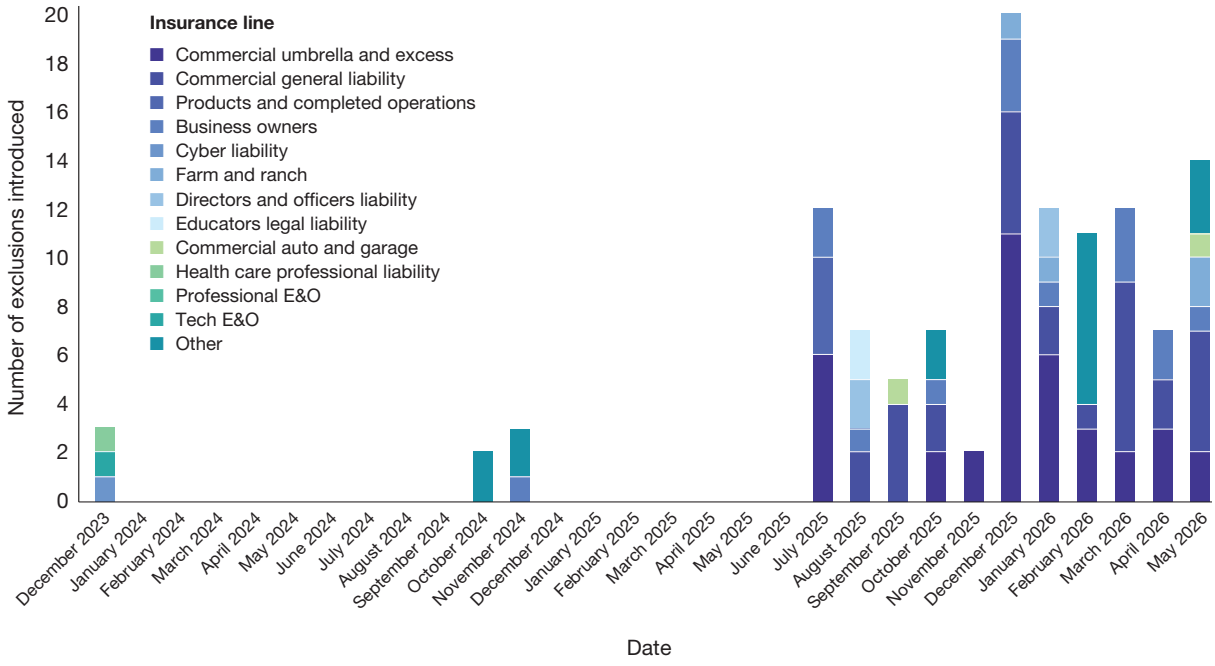
SOURCE: Reproduced from Berkley Insurance Company, 2024.

losses because of "a flawed AI-generated structural calculation, a hallucinated building specification, or an erroneous project schedule [leading] to design defects, construction delays, accidents, and resulting bodily injury and/or property damage claims" (LaMorte, 2026).

Some Carriers are Affirmatively Covering AI Losses

While the major carriers mentioned earlier exclude AI-related losses, a small number of carriers are affirmatively covering these risks. Overall, these policies cover harms or losses caused by AI outputs or failures in AI system behavior, such as hallucinations or erroneous outputs, underperformance and technical devia-

FIGURE 4.2
AI Exclusions in Admitted Policies



SOURCE: Features information from the data collected from FilingFocus, undated.

NOTE: Carriers file exclusions in each state in which they operate. These results are based on searches for *artificial intelligence* and therefore may not be able to distinguish between GenAI and machine learning. However, we infer that filings submitted after 2023 refer to GenAI.

tions from expected behavior, IP and data-related harms, and third-party legal liability for those harmed by AI outputs (including bodily injury and property damage). For example, Munich RE covers first- and third-party AI losses suffered by deployers (users) of AI tools or by AI developers (AI labs) and vendors (building AI tools) for losses relating to the following (Munich RE, undated):

- underperformance of the models
- hallucination, misleading content and false information
- bias and fairness risks, leading to discrimination
- privacy infringement following the leakage of private or sensitive information
- IP violations by models trained on copyright-protected material
- users and customers exposed to harmful AI-generated content.

Other traditional carriers cover similar AI-related losses. For example, in late 2024, AXA XL expanded coverage to include data poisoning, regulatory fines or fees, and lawsuits arising from copyright or use rights violations (AXA, 2024).[3] Coalition, a cyber insurance managing general agent, revised its cyber insurance policy to cover losses resulting from AI system failures and AI-enabled fraudulent fund transfers, including deepfakes (Coalition, 2024). CFC, a specialty cyber carrier, is addressing AI risk in a number of its core policies, including "Technology Errors & Omissions (Tech E&O), Professional Liability (E&O), eHealth, Intellectual Property (IP), Management Liability, Media, and Cyber Proactive Response (CPR)" (Jimenez-Sanchez, 2026).

A number of new insurance providers have also emerged to cover AI-related losses exclusively. For example, Testudo offers a stand-alone specialty liability policy with limits up to $10 million per policyholder tar-

[3] *Data poisoning* refers to injecting a model's training data with false, outdated, or incomplete data.

geted for *deployers* of GenAI tools (rather than model developers) (Lewin-Smith and Titmarsh, 2026). Testudo only covers third-party liability arising from a range of harms caused by AI outputs, including

- IP infringement and defamation claims, including libel, slander
- hallucinations leading to financial loss
- AI regulatory violation
- bodily injury resulting from reliance on, or interaction with, AI-generated outputs
- unauthorized data disclosure via AI-generated outputs.

Testudo's approach to evaluating AI risk primarily relies on the company's proprietary litigation dataset and other risk signals, which enables "real-time risk scores" for AI models and uses by their policyholders (Lewin-Smith and Titmarsh, 2026).

Armilla offers primary and excess coverage for third-party liability claims and associated costs arising from AI model underperformance affecting both developers and deployers of AI systems, including GenAI and AI agents. Coverage is triggered by specified measures of model underperformance, which may include hallucinations, model errors, nonbreach data leakage, AI regulatory violations, AI-related IP or copyright infringement, errors in AI agent decisions or execution, and other departures from expected AI behavior (Armilla, 2025b; Mixides, 2025).

Armilla also provides (first-party) AI performance guarantees for developers of GenAI and AI agent tools. These policies allow AI vendors to insure their contractual assurances about system performance, such as agent decisions and hallucination rates, and pay the customer if the AI tool fails to satisfy specified performance indicators. However, there are still exclusions for this warranty product related to misuse of the AI system by users, unauthorized changes by the policyholder, or events beyond the vendor's control. Rather than offering an enterprisewide policy, Armilla's structures its coverage per model, with limits up to $5 million per model use and $25 million per policyholder (Armilla, 2025a).

The Artificial Intelligence Underwriting Company (AIUC) offers an AI insurance policy that is functionally similar to a tech E&O policy (see AIUC, undated) and provides up to $50 million in coverage for both first-party and third-party AI losses. In addition, unlike most other AI policies, it also affirmatively covers losses caused by AI *agents,* such as tool call failures; financial losses from incorrect payments or refunds; and brand and reputational harm resulting from harmful or inappropriate AI outputs.

AIUC's underwriting approach centers on the development of an AI certification, which examines and audits an AI model's compliance around security, safety, reliability, privacy, accountability, and societal risks (see Standard 1 [AIUC-1], undated-a). AIUC evaluates compliance with the certification through initial technical testing and annual reviews, supplemented by quarterly technical assessments.

Policies also vary in their AI-specific coordination clauses from the *other insurance* provision, which coordinates coverage when multiple policies apply to the same loss and specifies the division of payment responsibility among insurers (International Risk Management Institute, undated). In practice, this provision determines whether the AI policy responds on a primary, excess, or shared basis relative to other available insurance. For example, AIUC's policy responds on a primary basis, while Testudo responds on an excess basis.

> Most AI policies still remain silent on AI agents, while AIUC and Armilla affirmatively cover them.

In addition, even stand-alone AI policies seek to limit coverage for some AI-related losses through AI-specific exclusions. For example, Relm Insurance has two notable exclusions that relate to intentional acts and to losses from known model misbehaviors:

Intentional Acts and Fraud

Knowingly using unauthorized training data, intentionally violating IP rights, or committing fraud are excluded under standard policies. If the insurer can show you knew about the infringement or violation, coverage may be denied.

Prior Knowledge and Known Circumstances

A known circumstance is a problem you were aware of before the policy started, and insurers often exclude claims arising from it. If you knew about a model defect before policy inception and didn't disclose it, claims arising from that defect may be excluded. (Relm Insurance, undated)

The Remaining Carriers Are Silent on AI Coverage

The approaches we have described reflect differing carrier views about the insurability of AI risk. Some incumbents are moving to broadly exclude AI-related losses because the exposure is diffuse, difficult to price, hard to disentangle from other perils, and potentially correlated across multiple lines. At the same time, other carriers see a significant commercial opportunity in what may become a large new book of business and are positioning themselves to lead the market through affirmative AI coverage.[4]

The remaining carriers (and, indeed, potentially the same carriers for different lines) are taking a third approach: remaining silent on the risk. In these cases, policies neither expressly affirm nor expressly exclude AI-related losses. From the perspective of some policyholders, this silence does not necessarily create ambiguity. If AI is not excluded and if the alleged loss otherwise falls within the policy's insuring agreement, policyholders may reasonably expect the claim to be covered. More broadly, many policyholders operate on the assumption that they remain covered under traditional lines unless and until insurers adopt explicit AI exclusions.

At the same time, policyholders should not equate silence with settled coverage. Whether an AI-related loss is covered still depends on the policy language; the theory of liability; the characterization of the loss; and the interaction of existing exclusions, conditions, and definitions that were not drafted with AI in mind. Silence may therefore preserve the possibility of coverage but also increases the likelihood of dispute, inconsistent treatment across carriers, and uncertainty about where AI risk ultimately resides.

This posture is consistent with how insurance markets have historically responded to emerging risks and closely resembles the earlier evolution of cyber insurance (Woods and Moore, 2020).

Given this discussion, the most immediate consequence for insurers and policyholders is that AI-related coverage can no longer be inferred from the mere existence of traditional cyber, liability, or professional lines but increasingly depends on the specific policy language. Two firms using similar AI systems may therefore face very different risk outcomes depending on how their policies cover, exclude, or remain silent on AI losses. The coexistence of affirmative cover, exclusions, and silence means that AI risk is being allocated unevenly across insurers and policyholders. This is not necessarily a problem, given that different carriers choose to address the risk differently. However, it does suggest a heightened need for policyholders and carriers to establish a clear and common understanding of losses that are covered those that are not.

[4] Our analysis of the policy data (from FilingFocus, undated) helped identify carriers that exclude AI-related losses. We also know that no stand-alone AI policies have been filed in the admitted market as of this writing. By inference, therefore, we assume that the remaining carriers (and their policies) stand silent on AI coverage.

CHAPTER 5

Underwriting AI Risk Requires Combining Imperfect Tools

The uneven market responses we have described (broad exclusions from some carriers, emerging affirmative products from others, and silence from most) reflect, in part, the difficulty of assessing AI risk in the first place. Without mature claims data or standardized evaluation methods, carriers have little consensus on how to price AI exposure or set appropriate limits. Therefore, in this chapter, we leverage insights from interviews and industry literature to examine the principal methods carriers can use to assess (and ultimately price) AI-related risk in the absence of a mature loss record. We focus on four main inputs: historical claims and incident data, quantitative and qualitative model evaluation, and scenario analysis.

Historical claims remain the most useful underwriting input because they provide direct evidence of loss frequency, severity, and variance. When available, AI-related claims identify recurring failure modes and help carriers compare loss distributions (e.g., frequency and severity) across AI use cases. A sufficiently large number of claims and data derived from claims can also help identify which governance and security controls are most able to prevent loss (Kenneally, 2021). For example, by merging insurance policy application data (i.e., information about which risk-mitigating controls the policyholder has implemented) with claims data (i.e., type and severity of a claim), standard econometric techniques help explain the correlation and marginal contribution of these risk mitigation techniques and the probability of filing a claim (Romanosky, 2022).

> Approaches for assessing AI risk rely on traditional approaches, such as historical claims, incident data, quantitative and qualitative methods, and scenario analysis.

The difficulty is that actual AI claims are only starting to emerge, so any inferences are highly unstable. The underlying models, deployment contexts, and legal theories of loss are also changing quickly, further limiting the predictive value of past claims. In addition, there is an unavoidable dynamic with AI models (possibly more so than with traditional software) that the *same* model configured the *same* way may generate very different failures (and therefore different loss outcomes). For these reasons, current historical claims alone are unlikely to provide a sufficient actuarial basis.

As with claims data, **incident data** can provide a broad view of how AI systems fail in practice. Relevant examples (as examined earlier) include actual AI incidents, lawsuits, near misses, and regulatory enforcement. Actual incidents and near misses can help identify first- and third-party loss pathways, such as business interruption, data leakage, or model errors. Naturally, these data are most useful when collected and classified transparently and consistently, with enough detail to evaluate the underlying root cause, use case, consequence, and any controls that were (or were not) in place at the time of an incident (see Trout et al., 2026, Appendix 1). Not all incidents, of course, will materialize into insurable loss events.

Lawsuits, regulatory enforcement, and settlements are useful for identifying (third-party) liability pathways, legal theories of loss, and emerging areas of claim activity. However, months or years may elapse

between the alleged wrongful act or occurrence and formal disposition, reducing usefulness of these data for insurance purposes.

Quantitative and qualitative evaluation techniques can also be applied to AI systems. Quantitative AI evaluations seek to estimate model error rates and failure probabilities. For example, Liang et al. (2022) developed a testing framework that evaluates LLMs across seven metrics: accuracy, calibration, robustness, fairness, bias, toxicity, and efficiency. Chiang et al. (2024) developed a testing platform specifically for chatbots. Another highly quantitative approach is *conformal prediction*, which estimates the probability that model outputs will fall outside an acceptable tolerance (Singh et al., 2024).

In practice, these methods are most useful when the task can be reduced to a measurable benchmark with a defensible ground truth, for example, a customer-service chatbot answering questions about a retailer's (known and verifiable) list of products or return policies. Open-ended tasks, such as employment decisions or broad strategic advice, are more complicated with these approaches because errors are harder to define and verify. Additionally, it is difficult to explore the full range of outputs from a generative model; therefore, a measured error rate may not capture the rare failures that matter most for insurers.

Qualitative assessments collect structured evaluations of the insured's governance and control environment. These assessments may capture use cases and criticality of the deployed system, human oversight, model provenance, validation practices, security and privacy controls, vendor dependence, update management, documentation, and incident response. Examples of such qualitative efforts come from AIUC-1, Appia Foundation, the International Organization for Standardization's International Electrotechnical Commission, the National Institute of Standards and Technology's AI Risk Management Framework, Cloud Security, and Alliance's AI Controls Matrix (see AIUC-1, undated-b; Appia Foundation, undated; International Organization for Standardization, International Electrotechnical Commission 42001, 2023; National Institute of Standards and Technology, undated; Cloud Security Alliance, 2026).

Such qualitative reviews can be applied across a wider range of AI deployments, relative to formal error modeling, but rely more heavily on judgment and self-reported information. In practice, however, carriers are likely to use both quantitative and qualitative methods to capture as much information as possible.

Scenario analysis is a common forward-looking underwriting tool for identifying plausible loss pathways in the absence of (or together with) mature claims records (International Actuarial Association, 2013). Carriers use technical understanding, plausible use cases, and available historical evidence (such as incident and loss events) to develop and prioritize loss scenarios. For example, Lloyd's of London conducts scenario analysis for natural and human-made risks annually (Lloyd's, 2024; Lloyd's, 2026) and performed a similar analysis for AI loss scenarios. To be clear, these are not catastrophic scenarios that address AI losses in, for example, autonomous vehicles, AI-enabled drinking water systems, space, crime, medical systems, and marine systems (Powel, 2025). These exercises are useful for assessing the scope and potential *severity* of loss rather than estimating event *frequency*. Other efforts, by CyberCube and AIUC, develop AI scenarios to the extent that AI can be an accelerator for cyber activity.

Taken together, these underwriting tools are individually limited and collectively immature. Carriers cannot yet reliably estimate the frequency or severity of AI losses, which makes pricing uncertain and explains much of the market divergence seen in Chapter 4. In practice, this uncertainty is likely to encourage conservative underwriting responses, including narrower terms, higher retentions, lower limits, selective affirmative coverage, and broader exclusions. The limitation becomes especially consequential when insurers consider accumulation risk, because uncertainty about individual AI losses can also obscure correlated exposure across insureds, vendors, infrastructure providers, and insurance lines.

CHAPTER 6

Five Mechanisms That Could Produce AI Accumulation Risk for Insurers

Next, we identify mechanisms in which AI use could lead to accumulation risk (also known as *correlated failure* or *systemic risk*). This form of accumulation risk can be understood as a single event generating losses across many firms, industries, lines of insurance, or jurisdictions. The relevant feature is that the harm is correlated across insureds because they share the same dependency, exposure, or liability trigger. For the purpose of this research, we did not consider extreme forms of catastrophic failures that could rise to the level of a global pandemic or financial collapse but, rather, a scope of impact that affects many firms and many insurance carriers.

Universal AI attacks on AI systems are analogous to the problem in cyber risk in which many policyholders operate the same software or hardware that suffers from a common vulnerability, which attackers then exploit at scale, such as in the cases of the Code Red, Melissa, Conficker, and NotPetya malware (Vaknin, 2013; Steinberg, Stepan, and Neary, 2021). With AI, adversaries could orchestrate scaled attacks against customer service or finance AI agents to authorize payments, sign legal contracts, or disclose sensitive information. For example, a *prompt injection* is a specific kind of security control bypass attack that tricks the model into revealing sensitive or unsafe information or into executing inappropriate instructions (such as, "ignore previous instructions and do [this malicious thing]"). Typically, these cleverly crafted prompts work on individual models and versions; however, *universal prompt injections* are especially dangerous because they work across many different models and vendors (Liu et al., 2024; McCauley et al., 2026). Other kinds of universal attacks have been demonstrated against multimodal (e.g., text and image) AI models (Rahmatullaev et al., 2025).

Accumulation risk can also occur from **common model dependency** when firms rely on the same underlying model, component, or infrastructure (Bommasani et al., 2022). If that model contains a latent defect (e.g., it was trained from biased or poisoned data) or receives a flawed update, losses may emerge across many downstream firms. Cyber-related examples include the SolarWinds compromise (U.S. Government Accountability Office, 2021) and the corrupted CrowdStrike update that caused devices running the Windows operating system to fail on boot (Kerner, 2024). Furthermore, the AI technology stack is highly concentrated at both the hardware and software layers. At the manufacturing level, Taiwan Semiconductor Manufacturing Company produces approximately 90 percent of the world's most-advanced computing chips, while NVIDIA controls an estimated 80 to 90 percent of the AI hardware market by revenue (as well as the software that interfaces with the hardware) (Cloud Security Alliance, 2026). Moreover, many of these AI services run within massive hyperscaler (data center) facilities, which are concentrated across only three main vendors: Amazon Web Services, Microsoft Azure, and Google Cloud (Synergy, 2025).

AI can also serve as a **force multiplier for harmful activity**. In this case, the AI system is not itself the common point of failure but, rather, a tool that amplifies the speed, scale, or success of malicious attacks and expands attacker capability in ways that create more correlated losses than would otherwise occur or be possible. For example, concern is growing about the rise in cyberattacks powered by AI-discovered vulnerabili-

ties (Snape, 2026). Extreme examples include the development of AI-enabled biological, chemical, or kinetic weapons that could produce widespread physical harm (Hersman, Blackwelder, and Parachini, 2026). While human-powered cyberattacks may still represent the vast majority of cyber events, at least for phishing and social engineering (Cartier, 2026), there is a legitimate concern about increasing AI use across all stages of a cyberattack. For example, announcements in April 2026 about attackers using AI tools to more readily identify and develop exploit code for software vulnerabilities raise even more concern about scalable AI-powered cyberattacks (Evans, 2026; Google Threat Intelligence Unit, 2026).

AI accumulation risk may also arise from a **legal or regulatory shock** of a common practice, as well as from compliance with that practice. A new law, court ruling, or agency enforcement could make previously widespread AI use actionable across many firms at once, such as AI-hiring or AI-lending model use or AI-generated content leading to claims of copyright infringement. Moreover, consider a new law requiring firms to disclose AI use to customers. Noncompliance with such a law could trigger regulatory enforcement or could trigger securities litigation alleging that officers failed to oversee compliance of the tool, misrepresented the firm's AI practices, or omitted material information from public disclosures. Here, the same legal development could generate claims across multiple insureds at roughly the same time. Depending on the particular law or regulation, losses may also emerge across different lines of insurance, further obscuring the accumulation costs. This mechanism is especially relevant when there is substantial latency between an AI incident and the resulting harm. If a widespread practice causes injury that is only discovered years later, a change in legal doctrine or regulatory enforcement may activate claims across multiple historical policy years at once, potentially straining insurer reserves, particularly under occurrence-based liability coverage.

A final possible behavior that may induce correlated loss from AI systems broadly, although it is much more amorphous, is that losses may arise from a slow, evolving, and **subtle corruption or degradation** of model performance (i.e., model drift[1]) or may output such behavior as *AI sycophancy* (the tendency for LLMs, such as ChatGPT, to excessively flatter or agree with the user), lying, or deception (Apollo Research, 2026) that leads to insurable losses. Unlike the previous classes of aggregation, these events may not have a single triggering event (such as a malicious attack or corrupt model update) but simply manifest as first- or third-party claims across many insureds, policy lines, and carriers. Given the relatively sudden introduction of large models, and their rapid development, this category is more speculative but deserves attention and further research.

Taken together, these mechanisms suggest that AI accumulation risk can plausibly arise from a number of root causes. For insurers and reinsurers, the principal challenge (as always) is that AI-related losses may be correlated even when they do not appear identical at the claim level.[2] As a result, traditional line-by-line exposure management may understate AI-related accumulation. Managing this risk will require dependency-aware scenario analysis, root cause monitoring across insurance lines, and clearer visibility into when AI-related losses are covered, excluded, or left unresolved.

---

[1] As mentioned, *model drift* refers to a gradual loss of an AI model's predictive accuracy over time because of changes in the real-world environment and incoming data since the model was trained.

[2] It is possible that emergent behavior from AI agents may also materialize to create accumulation risk. However, as of this writing, these behaviors have not appeared. See Trout et al. (2026) for discussion of multiagentic failure.

CHAPTER 7

Conclusion and Recommendations

Conclusion

Given the surge in deployment of AI technologies, it seems clear that the insurance and reinsurance markets will grow to cover the inevitable losses and liability (Abraham and Sharkey, 2024). However, as AI harms occur with greater frequency and severity, two important policy metrics become relevant: the size of the AI protection gap and the extent of accumulation risk. The protection gap represents the difference between economic losses and the insured proportion of loss (Schanz, undated). While it is neither necessary nor socially efficient for the gap to reach zero,[1] a smaller gap is generally a signal of economic resilience (Schernberg, 2026).

Tracking these two metrics is critical for insurance stakeholders and regulators because the metrics speak to the fragility and magnitude of this emerging risk and are also linked through carrier management of AI insurance, as shown in Figure 7.1.[2]

When more AI loss is *affirmatively covered*, the protection gap shrinks as firms transfer risk to insurance. However, this comes at a cost for carriers because they absorb a greater share of accumulation exposure. On the other hand, when carriers *exclude* AI-related losses, insurer-side accumulation exposure declines but at the cost of increasing the protection gap as more of that risk remains with firms and the broader economy. *Silent* AI coverage is the most problematic because it obscures both the size of the protection gap and accumulation exposure, making it difficult for policyholders, insurers, and regulators to determine what is insured, what is retained, and where correlated losses may emerge.

Recommendations

For these reasons, we offer three recommendations so that insurers and regulators can better measure and manage the protection gap and accumulation exposure.

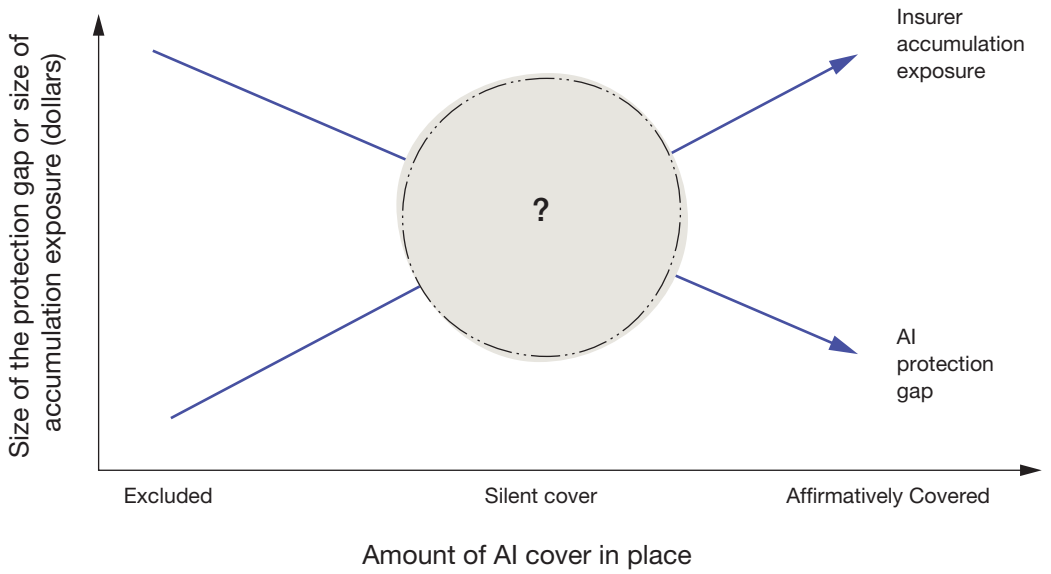
Improve Comparability and Transparency Across Differing Insurance Policies by Adopting a Standardized Notice Structure

We have shown that AI harms do not map neatly onto a single line or consistently across insurance lines. Moreover, we have also shown that carriers are diverging between affirmative cover driven by a few incumbent carriers, broad exclusions driven in part from a surge in admitted market filings, or are remaining silent. To some extent, startup AI companies help resolve the gap that would be created by exclusions providing

[1] The notion of a protection gap of zero suggests insurance cover for 100 percent of a loss, which would not be socially efficient, because insurance is costly, and not every loss needs to be insured. Furthermore, we note that the protection gap is largely driven by frictions in the market rather than true social inefficiency.

[2] The figure implies a trade-off between AI cover and accumulation exposure. However, we recognize that not every AI risk may produce accumulation risk (i.e., idiosyncratic model errors) because they could be diversified across policyholders.

FIGURE 7.1
Relationship Between AI Cover, Insurer Accumulation Exposure, and Protection Gap



NOTE: The figure illustrates the change in the size of the AI protection gap and accumulation exposure (in dollars) as more AI losses are affirmatively covered. Unfortunately, as of mid-2026, we know of no available data that could be used to estimate or approximate the size of either.

AI-risk cover. Nevertheless, these dynamics create ambiguity and confusion for policyholders about which policy responds first, whether more than one responds, whether one policy excludes what another assumes is covered, or whether the loss falls into a gap between lines.

Therefore, state insurance regulators and the NAIC should develop an AI Coverage Notice that enables carriers to clarify how and under which lines AI-related harms are affirmatively covered, excluded, or left silent.[3] The coverage notice would specify covered AI uses (if at all), triggers, any conditions around "other" insurance, and resolution criteria to manage overlapping responding lines. State regulators and the NAIC need not require that carriers insure AI the same way or even have a common understanding of the risk but, rather, should ensure that the structure and requirements of the coverage notices are consistent across carriers. In addition, policy documentation of AI use should be clarified across predictive AI, GenAI, and agentic AI.

We recognize that carriers may resist standardized notice structures if the carriers believe AI risk differs materially across lines, products, and policyholder uses or view additional disclosure as burdensome or litigation-sensitive. The goal, therefore, is not to impose uniform coverage but to improve comparability and transparency across differing coverage approaches. We believe these concerns can be mitigated by standard-

[3] We deliberately focus on state regulators and industry-led governance rather than federal intervention. In our assessment, additional legislation or regulation at the federal level is not warranted. Moreover, the regulation of the business of insurance has historically resided with the states rather than the federal government, under the McCarran-Ferguson Act (U.S. Code, Title 15, §§ 1011–1015). We acknowledge, however, that this allocation of authority is now subject to renewed federal attention. Shortly before this writing, the White House issued Executive Order 14365, which directs federal action to constrain state AI laws deemed to conflict with national policy and to prepare legislative recommendations preempting such laws (Executive Order 14365, 2025). In response, the NAIC urged the administration to "at a minimum, affirm state regulation of AI in the business of insurance to avoid damaging uncertainty" (NAIC, 2025). This matter has yet to be resolved as of this writing.

izing only the notice format and core definitions, while allowing carriers to retain flexibility over substantive coverage terms and line-specific wording.

Standardize Data Collection Taxonomy for Attritional and Accumulation Loss

Harms from AI *incidents* generally materialize quickly but do not always convert into insurable losses. On the other hand, AI *lawsuits* are more informative of costs but are quite lagged in time. Moreover, public information on AI claims data is nonexistent, and insurers' AI assessment approaches are heterogeneous at best. The consequence is that insurers lack the necessary information needed to price AI risk appropriately (even within their own book) or to effectively compare AI use cases. Carriers' ability to estimate which controls or practices most reduce AI risk is also highly constrained, which is also a chronic problem for cyber insurance.[4]

Therefore, policy researchers and brokers, carriers, and reinsurers should develop a common taxonomy for AI incidents, claims, and controls. The taxonomy would form the basis for a reporting infrastructure that translates AI failures into insurable events and creates an empirical basis for better understanding the frequency and impact of AI harms across sectors and use cases. The taxonomy would, to the extent possible, capture the technical event, harm caused, economic consequence, claim or legal action, coverage line(s) implicated, and controls and governance conditions at time of loss.

Carriers can use the data schema to privately track losses resulting from AI systems, thereby improve underwriting. Each carrier's data would be treated as confidential and proprietary, with no expectation that the data would be shared across any stakeholders, of course. This standardization also facilitates aggregated reporting to regulators in cases of severe AI incidents, if this becomes warranted.

The main obstacles here are definitional, operational, and competitive. AI incidents, claims, controls, and loss pathways are still evolving, which makes it difficult to design a taxonomy that is both durable and flexible. Carriers and brokers may also hesitate to adopt common reporting structures if they fear disclosure of proprietary underwriting methods or commercially sensitive claims information. One way to solve this is to use a small common schema tied to insurable events, allowing firms to retain detailed internal data privately, and limiting any shared reporting to aggregated or anonymized outputs.

Develop and Analyze Accumulation Scenarios

AI accumulation risk may arise from common model dependencies, shared cloud or hardware infrastructure, universal attacks, legal or regulatory shocks, and other correlated failure modes. Even if these risks are not yet comparable to catastrophic cyber or terrorism, they are important to track.[5]

Therefore, insurers and reinsurers should conduct and disclose AI aggregation scenario analysis focusing on actual deployment use cases of AI tools, shared model dependency, common infrastructure failure, and regulatory shock. These exercises will help establish current thinking regarding modes and consequences of aggregation failure and promote the foundation for continued research. State insurance regulators and NAIC can support this effort by collecting and reporting annual anonymized incident reports.

This recommendation also involves implementation frictions. Many insurers may lack visibility into upstream AI dependencies, especially when policyholders rely on third-party models, application programming interfaces, cloud services, or embedded tools that are not fully disclosed. Scenario analysis can also

[4] For a review of the difficulty carriers have identifying and incentivizing cyber risk reduction, see Schwarcz and Wolff (2026).

[5] We recognize that there are avenues for AI to cause both nationally and globally catastrophic harms from AI-discovered pandemics or other human-extinction level events. However, these are beyond the scope of U.S. insurability and, therefore, of this research.

become resource-intensive and hard to compare across firms unless it is built around a small set of common scenarios and materiality thresholds. These barriers can be reduced by focusing only on material AI use, adopting a fixed scenario set, and using annual aggregate reporting with ad hoc updates only when major dependency or legal changes occur.

In Summary

We believe AI insurance will become more than a niche specialty product. As they work to clarify coverage, improve incident reporting, and monitor accumulation risk, insurers, regulators, and policyholders will also have the potential to shape the practical standards for what counts as insurable and, therefore, acceptable AI use. The challenge lies in whether the insurance industry can adapt quickly enough to keep pace with the technology's deployment. As AI becomes more embedded in critical business and social functions, the need for clearer coverage, better underwriting data, and more systematic accumulation monitoring will only grow.

APPENDIX

Insurance Lines

Table A.1 describes the relevant insurance lines and the kinds of losses the lines cover.

TABLE A.1
Descriptions of Insurance

Insurance Line	Typical Losses Covered
Media liability	Claims arising from published or distributed content, including defamation (libel, slander), invasion of privacy, copyright or trademark infringement in media, plagiarism, and related legal defense costs
Technology E&O	Customer financial losses that are due to failures in technology products or services, such as software bugs, implementation errors, system outages, negligent professional services, or failure to perform as contracted
Miscellaneous professional liability	Professional negligence claims against service providers outside traditional professions (e.g., consultants, recruiters, marketing firms) that cover economic losses resulting from errors, omissions, or failure to deliver professional services
Professional indemnity	Similar to E&O (primarily used outside North America); covers liability for negligent professional advice or services that cause clients financial loss, including legal defense costs
Commercial general liability	Third-party claims for bodily injury, property damage, and personal or advertising injury arising from business operations; typically excludes purely economic losses and most cyber-related losses
IP	Defense costs and, depending on the policy, damages arising from allegations of IP infringement (e.g., patent, copyright, trademark, trade dress) or enforcement costs for the insured's own IP rights; coverage varies substantially by policy
Cyber	First-party and third-party losses arising from cyber incidents, including data breaches, ransomware, business interruption, incident response, digital forensics, notification costs, regulatory investigations, cyber extortion, privacy liability, and network security liability
Crime	Losses from fraud or dishonesty, such as employee theft, computer fraud, funds transfer fraud, social engineering (if endorsed), forgery, embezzlement, and theft of money or securities
Directors and officers liability	Claims alleging wrongful acts by directors and officers, including breach of fiduciary duty, mismanagement, securities claims, shareholder lawsuits, and regulatory investigations; covers defense costs and settlements, subject to policy terms
Employment practices liability	Claims by employees or applicants alleging wrongful employment practices, including discrimination, harassment, wrongful termination, retaliation, failure to promote, and certain wage-and-hour defense costs, where endorsed

Abbreviations

AI	artificial intelligence
AIIDB	Artificial Intelligence Incident Database
AIUC	Artificial Intelligence Underwriting Company
AIUC-1	Artificial Intelligence Underwriting Company Standard 1
CSAM	child sexual abuse material
DAIL	Database of AI Litigation
E&O	errors and omissions
GenAI	generative artificial intelligence
GPT	Generative Pre-trained Transformer
IP	intellectual property
LLM	large language model
NAIC	National Association of Insurance Commissioners
SERFF	System for Electronic Rate and Form Filing
URL	Uniform Resource Locator

References

Abraham, Kenneth S., and Catherine M. Sharkey, "The Glaring Gap in Tort Theory," *Yale Law Journal*, May 31, 2024. As of August 8, 2026:
https://yalelawjournal.org/article/the-glaring-gap-in-tort-theory

Acharya, Deepak Bhaskar, Karthigeyan Kuppan, and B. Divya, "Agentic AI: Autonomous Intelligence for Complex Goals—A Comprehensive Survey," *IEEE Access*, Vol. 13, January 2025.

AIAAIC, "AIAAIC Repository," webpage, 2026. As of August 17, 2026:
https://www.aiaaic.org/aiaaic-repository

AIUC—*See* Artificial Intelligence Underwriting Company.

AIUC-1—*See* Artificial Intelligence Underwriting Company Standard 1.

AI Vulnerability Database, homepage, undated. As of August 17, 2026:
https://avidml.org/

AON, "Artificial Intelligence Risk Management in 2025," fact sheet, 2025.

Appia Foundation, "Translating Global Standards into Assessable Criteria for the AI Value Chain," webpage, undated. As of August 17, 2026:
https://appiafoundation.org/

Armilla, "AI Liability Insurance Fact Sheet," 2025a.

Armilla, "Chaucer and Armilla Launch New AI Liability Insurance Product," press release, April 23, 2025b. As of August 18, 2026:
https://www.armilla.ai/resources/chaucer-and-armilla-launch-new-ai-liability-insurance-product

Artificial Intelligence Underwriting Company, "The World Is Not Ready for the Risk AI Agents Bring," webpage, undated. As of August 17, 2026:
https://aiuc.com/product

Artificial Intelligence Underwriting Company Standard 1, "Frequently Asked Questions," webpage, undated-a. As of August 17, 2026:
https://www.aiuc-1.com/faq

Artificial Intelligence Underwriting Company Standard 1, "The Standard for AI Agent Security, Safety and Reliability, webpage, undated-b. As of August 17, 2026:
https://www.aiuc-1.com/

Apollo Research, "We Need a Science of Scheming," webpage, January 19, 2026. As of May 8, 2026:
https://www.apolloresearch.ai/science/science-of-scheming

AXA, "AXA XL Unveils New Cyber Insurance Extending Coverage to Help Businesses Manage Emerging Gen AI Risks," press release, October 21, 2024.

Berkley Insurance Company, "Artificial Intelligence Exclusion," form, June 2024.

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al., "On the Opportunities and Risks of Foundation Models," arXiv, arxiv.org/abs/2108.07258, last revised July 12, 2022.

Braun, Virginia, and Victoria Clarke, "Using Thematic Analysis in Psychology," *Qualitative Research in Psychology*, Vol. 3, No. 2, 2006.

Brew, Oliver, and Eden Fall-Bailey, "Ready or Not: The Impact of Artificial Intelligence on Insurance Risks," Lockton Re and Armilla, 2026.

Carter, Tom, "A Founder Says Cursor's AI Agent Deleted His Startup's Database, Causing Chaos for Customers," *Business Insider*, April 28, 2026.

Cartier, Maxime, "AI Phishing Attacks: How Big Is the Threat? (+Infographic)," Hoxhunt, updated July 16, 2026. As of August 18, 2026:
https://hoxhunt.com/blog/ai-phishing-attacks

Chiang, Wei-Lin, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Hao Zhang, Banghua Zhu, Michael Jordan, Joseph E. Gonzalez, and Ion Stoica, "Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference," arXiv, arxiv.org/abs/2403.04132, March 7, 2024.

Cloud Security Alliance, "AI Controls Matrix," v.1.1, webpage, June 22, 2026. As of August 17, 2026: https://cloudsecurityalliance.org/artifacts/ai-controls-matrix-v1-1

Coalition, "Coalition Adds New Affirmative AI Endorsement to Cyber Policies," press release, March 26, 2024.

Coalition, "2026 Cyber Claims Report," webpage, 2026. As of August 18, 2026: https://www.coalitioninc.com/claims-report/2026

"Deepfake Images of Taylor Swift Spread on Social Media. Her Fans Are Fighting Back," Associated Press, January 26, 2024. As of June 30, 2026:
https://apnews.com/article/
taylor-swift-ai-images-protecttaylorswift-nonconsensual-d5eb3f98084bcbb670a185f7aeec78b1

Doe v. X.AI Corp., 5:26-cv-00772 (N.D. Cal. 2026).

Dolly v. GitLab Inc., No. 5:24-cv-06244 (N.D. Cal. 2024).

Ethical Tech Initiative, "DAIL—The Database of AI Litigation—Scope," webpage, George Washington University, undated. As of June 15, 2026.
https://blogs.gwu.edu/law-eti/dail-the-database-of-ai-litigation-scope/

Evans, Steve, "AI Treated as Force-Multiplier for Cyber Losses. Introduces Aggregation, Correlation Risks: CyberCube," webpage, Artemis, April 14, 2026. As of August 18, 2026:
https://www.artemis.bm/news/
ai-treated-as-force-multiplier-for-cyber-losses-introduces-aggregation-correlation-risks-cybercube

Executive Order 14365, "Ensuring a National Policy Framework for Artificial Intelligence," Executive Office of the President, December 11, 2025.

FilingFocus, homepage, database, undated. As of August 25, 2026:
https://filingfocus.com

Garcia v. Character Technologies, Inc., No. 6:24-cv-01903 (M.D. Fla. 2024).
https://www.courtlistener.com/docket/69300919/garcia-v-character-technologies-inc/

Germond, Nancy, "Verisk to Roll Out New General Liability Exclusions for Generative AI Exposures," press release, Big i, October 21, 2025.

Glaser, Barney G., and Anselm L. Strauss., *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Aldine Publishing Company, 1967.

Google Threat Intelligence Unit, "Adversaries Leverage AI for Vulnerability Exploitation, Augmented Operations, and Initial Access," *Google Cloud* blog, May 11, 2026.

Gross, Grant, "Insurance Carriers Quietly Back Away from Covering AI Outputs," CSO Online, April 16, 2026.

Goodman, Leo A., "Snowball Sampling," *Annals of Mathematical Statistics*, Vol. 32, No. 1, March 1961.

Guest, Greg, Arwen Bunce, and Laura Johnson, "How Many Interviews Are Enough? An Experiment with Data Saturation and Variability," *Field Methods*, Vol. 18, No. 1, February 2006.

Hartzog, Woodrow, and Daniel Solove, *Breached! Why Data Security Law Fails and How to Improve It*, Boston University School of Law, 2022.

Hersman, Rebecca K. C., Lauren Blackwelder, and John V. Parachini, *AI Implications for Chemical, Biological, Radiological, and Nuclear Defense Policy and Programs: A Strategic Analytic Research Agenda*, RAND Corporation, PE-A4611-1, April 2026. As of August 18, 2026:
https://www.rand.org/pubs/perspectives/PEA4611-1.html

Hill, Katrina, and Ed Pocock, "Smart Systems, Blind Spots: Rethinking Insurance for the AI Era," Gallagher Re, March 24, 2026.

Huang, Kalley, "Why Pope Francis Is the Star of A.I.-Generated Photos," *New York Times*, April 8, 2023.

Intelligence Council, "Berkshire, Chubb, and Travelers Are Removing AI Coverage. Substack.Com; P&C Insurance Executive Intelligence, April 27, 2026a. As of August 18, 2026: https://insuranceintel.substack.com/p/berkshire-chubb-and-travelers-are

Intelligence Council, "AI Liability: The $4.7B Specialty Insurance Opportunity," P&C Insurance Executive Intelligence, April 29, 2026b.

International Actuarial Association, *Stress Testing and Scenario Analysis*, 2013. As of August 18, 2026: https://actuaries.org/app/uploads/2025/04/StressTestingPaper.pdf

International Organization for Standardization, International Electrotechnical Commission 42001, "Information Technology—Artificial Intelligence—Management System," 2023.

International Risk Management Institute, "Other Insurance Clause," undated. As of August 18, 2026: https://www.irmi.com/term/insurance-definitions/other-insurance-clause

Jia, Ruo (Alex), Martin Eling, and Tianyang Wang, *Gen AI Risks for Businesses: Exploring the Role for Insurance*, Geneva Association, 2025.

Jimenez-Sanchez, Kassandra, "CFC Introducing Affirmative AI Coverage," Reinsurance News, June 26, 2026.

Jordan, M. I., and T. M. Mitchell, "Machine Learning: Trends, Perspectives, and Prospects," *Science*, Vol. 349, No. 6245, 2015.

Kenneally, Erin, "Hiding in Plain Sight: Towards Now-Gen Cyber Risk Underwriting," Guidewire, 2021. As of August 18, 2026: https://www.the-digital-insurer.com/wp-content/uploads/securepdfs/2021/09/1834-GuidewireCyenceRiskHidingInPlainSight.pdf

Kerner, Sean Michael, "CrowdStrike Outage Explained: What Caused It and What's Next," TechTarget, October 29, 2024.

LaMorte, Dylan, New Generative AI Insurance Exclusion: What the Construction Industry Needs to Know," *Construction Law Now Blog*, Cohen Seglias, April 20, 2026. As of August 18, 2026: https://www.cohenseglias.com/construction-law-now/new-generative-ai-insurance-exclusion-what-the-construction-industry-needs-to-know/

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton, "Deep Learning," *Nature*, Vol. 521, May 28, 2015.

Lewin-Smith, George, and Mark Titmarsh, "Testudo Launches New Insurance Coverage for Liability Risks Created by Generative AI Systems," press release, Testudo, August 1, 2026.

Liang, Percy, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, et al., "Holistic Evaluation of Language Models," arXiv, arxiv.org/abs/2211.09110, last updated October 1, 2023.

Liu, Xiaogeng, Zhiyuan Yu, Yizhe Zhang, Ning Zhang, and Chaowei Xiao, "Automatic and Universal Prompt Injection Attacks Against Large Language Models," arXiv, arxiv.org/abs/2403.04957, March 7, 2024.

Lloyd's, "Lloyd's 2024 Market Wide Scenario Exercise Report," December 2024. As of August 18, 2026: https://assets.lloyds.com/media/85c32cd8-440f-438f-951e-731b04c7dfcc/2024%20Lloyd's%20Market%20Exercise%20Report%20UCv.pdf

Lloyd's, "Annex 1—Full List of AI Scenarios," PDF, 2025.

Lloyd's, *RDS 2026: Realistic Disaster Scenarios Scenario Specification*, January 2026.

Lurye, Sharon, "How Scammers Are Using AI to Steal College Financial Aid," Associated Press, June 10, 2025.

Martin, Emmy, "Minnesota Solar Company Sues Google over False Information in AI Summary," *Minnesota Star Tribune*, June 13, 2025.

McCauley, Conor, Kenneth Yeung, Jason Martin, and Kasimir Schulz, "Novel Universal Bypass for All Major LLMs," April 24, 2025. As of August 20, 2026: https://www.hiddenlayer.com/research/novel-universal-bypass-for-all-major-llms

McGregor, Sean, "Preventing Repeated Real World AI Failures by Cataloging Incidents: The AI Incident Database," *Proceedings of the Thirty-Third Annual Conference on Innovative Applications of Artificial Intelligence*, 2021.

Mitchell, Tom M., *Machine Learning*, McGraw-Hill, March 1, 1997.

Mixides, Taylor, "Armilla Reveals Purpose-Built AI Liability Insurance amid Rising Legal and Regulatory Pressures," Reinsurance News, May 9, 2025.

Munich RE, "aiSure (TM) More AI Opportunity, Less AI Risk," webpage, undated. As of August 18, 2026: https://www.munichre.com/en/solutions/for-industry-clients/insure-ai.html

Munich RE, "Mind the Gap: A US-Focused Analysis of AI Liability Risks and the Implications for Insurance," 2024.

MIT AI Risk Initiative, homepage, undated. As of August 17, 2026: https://airisk.mit.edu/

NAIC—*See* National Association of Insurance Commissioners.

National Association of Insurance Commissioners, "Statement from the National Association of Insurance Commissioners (NAIC) on AI Executive Order," December 16, 2025.

National Institute of Standards and Technology, "AI Risk Management Framework," webpage, undated. As of August 18, 2026: https://www.nist.gov/itl/ai-risk-management-framework

The New York Times Company v. Microsoft Corporation, No. 1:23-cv-11195, S.D.N.Y., December 27, 2023.

Organisation for Economic Co-operation and Development Artificial Intelligence, "AIM: AI Incidents and Hazards Monitor," webpage, OECD.AI, 2026. As of August 17, 2026: https://oecd.ai/en/incidents

Orrick, "AI Law Center," webpage, undated-a. As of August 17, 2026: https://ai-law-center.orrick.com

Orrick, "U.S. AI Law Tracker—All States," dataset, Orrick AI Law Center, undated-b. As of June 30, 2026: https://ai-law-center.orrick.com/us-ai-law-tracker-see-all-states

Pain, Darren, and Sasha Romanosky, *Strengthening Cyber Resilience Through Insurance*, Geneva Association, 2026.

Polales, Horton & Leonardi, "If Your Company Uses GenAI, Be Aware of New Policy Exclusions for GenAI-Related Claims," February 10, 2026. As of August 20, 2026: https://phl-firm.com/generative-ai-insurance-exclusions-2026

Powel, David, "Understanding AI Exposures: AI Loss Scenarios Survey Results," Lloyd's Market Association, 2025.

Purifoy, Parker, "Lawyers Apologize for Fake AI Quotes in Trump Mass Layoffs Case," Bloomberg Law, May 16, 2026.

Rahmatullaev, Temurbek, Polina Druzhinina, Nikita Kurdiukov, Matvey Mikhalchuk, Andrey Kuznetsov, and Anton Razzhigaev, "Universal Adversarial Attack on Aligned Multimodal LLMs," arXiv, arxiv.org/abs/2502.07987, last revised June 4, 2025.

Raine v. OpenAI, Inc., No. CGC-25-628528 (Cal. Super. Ct., S.F. Cnty. 2025).

Relm Insurance, "Tech E&O for AI Products: What It Covers (Hallucinations, Model Errors, Bad Advice, IP and Training Data, and Contractual Liability)," undated.

Responsible AI Collaborative, *Founding Report*, March 28, 2022. As of August 19, 2026: https://asset.cloudinary.com/pai/cf01cce1af65f5fbb3d71fa092d001db

Romanosky, Sasha, "The State of Cybersecurity and Opportunities for Cyber Insurance," presentation, July 18, 2022. As of August 19, 2026: https://www.dol.gov/sites/dolgov/files/ebsa/pdf_files/2022-cybersecurity-insurance-and-employee-benefit-plans-romanosky-07-19.pdf

Roy, I., "Google Hit with Class Action Alleging Secret AI Email Scanning," Westlaw Data Privacy Daily Briefing, 2025 DPDBRF 4208, November 25, 2025. As of June 30, 2026: https://today.westlaw.com/Document/I05a690c8ca4d11f0bfdbcf11b1610f1f/View/FullText.html

Schanz, Kai-Uwe, "Understanding and Addressing Global Insurance Protection Gaps," research brief, the Geneva Association, undated. As of August 19, 2026:
https://www.genevaassociation.org/sites/default/files/research-topics-document-type/pdf_public/research_brief_-_global_insurance_protection_gaps.pdf

Schernberg, Hélène, *Addressing Growing Protection Gaps Through Better Public-Private Insurance Programmes*, Geneva Association, February 5, 2026.
https://www.genevaassociation.org/publication/public-policy-regulation/addressing-growing-protection-gaps-through-better-public

Schwarcz, Daniel, and Josephine Wolff, "The Limits of Regulating AI Safety Through Liability and Insurance: Lessons from Cybersecurity," *Fordham Law Review*, forthcoming.

Singh, Sukrita, Neeraj Sarna, Yuanyuan Li, Yang Lin, Agni Orfanoudaki, and Michael Berger, "Distribution-Free Risk Assessment of Regression-Based Machine Learning Algorithms," *Proceedings of the Thirteenth Symposium on Conformal and Probabilistic Prediction with Applications*, PMLR, Vol. 230, 2024.

Snape, Gia, "How Anthropic's Mythos Is Fueling Cyber Risk Aggregation Fears," *Insurance Business*, April 23, 2026.

Steinberg, Sean, Adam Stepan, and Kyle Neary, "NotPetya: A Columbia University Case Study," 2021.
https://www.sipa.columbia.edu/sites/default/files/2022-11/NotPetya%20Final.pdf

Stryker, Cole, and Mark Scapicchio, "What Is Generative AI?" IBM, undated.

Synergy, "Cloud Market Share Trends—Big Three Together Hold 63% while Oracle and the Neoclouds Inch Higher," Synergy Research Group, November 19, 2025.

Thele v. Google LLC, No. 5:25-cv-09704 (N.D. Cal. 2025).

Tsuki, Celestine Riza, "The Day Chevrolet's AI Chatbot Tried to Sell a $70,000 SUV for $1," webpage, Medium, August 12, 2025. As of August 19, 2026:
https://medium.com/@celestineriza/the-day-chevrolets-ai-chatbot-tried-to-sell-a-70-000-suv-for-1-29f4a1e954d9

Trout, Christian, Sanmi Koyejo, Sasha Romanosky, Giorgio Ripamonti, Lynn Thompson, Desiree Spain, Alex Taylor, Kevin Casey, Stephen Casper, Matthew Botvinick, et al., "Underwriting the Agent Economy: The Blueprint for an AI Insurance Stack," arXiv. arxiv.org/abs/2607.11999, last updated July 15, 2026.

United States v. Smith, No. 1:24-cr-00504 (S.D.N.Y. 2024).

U.S. Code, Title 15, Commerce and Trade; Chapter 20, Regulation of Insurance; §§ 1011–1015, McCarran-Ferguson Act.

U.S. Government Accountability Office, "SolarWinds Cyberattack Demands Significant Federal and Private-Sector Response," webpage, April 22, 2021. As of August 18, 2026:
https://www.gao.gov/blog/solarwinds-cyberattack-demands-significant-federal-and-private-sector-response-infographic

Vaknin, Avi, "The Top 5 Most Expensive Viruses and How to Avoid Them," webpage, Telx, December 11, 2013. As of August 19, 2026:
https://www.telxcomputers.com/computer-viruses-the-top-5-most-expensive-viruses-and-how-to-avoid-them/

Wells, Kate, "An Eating Disorders Chatbot Offered Dieting Advice, Raising Fears About AI in Health," NPR, June 9, 2023.

Woods, Daniel W., and Tyler Moore, "Does Insurance Have a Future in Governing Cybersecurity?" *IEEE Security & Privacy*, Vol. 18, No. 1, January–February 2020.

Zuidijk, Daniel, "Deepfakes in Slovakia Preview How AI Will Change the Face of Elections," Bloomberg, October 4, 2023.