SEPTEMBER 2026

The Insurance Industry's Retreat from AI Threatens to Slow Innovation and Adoption

By Gregory C. Allen

Executive Summary

On April 23, 2026, The Information **reported** that state insurance commissioners had quietly approved more than 80 percent of carrier requests to exclude AI-related damages from corporate insurance policies. That regulatory milestone attracted a fraction of the attention paid to this year's AI legislation, but it will likely matter more than most of it. For many, perhaps most, American businesses, an activity that cannot be insured is an activity that is effectively prohibited. Insurance has quietly become the most important de facto regulator of U.S. AI deployment, and it is currently regulating by withdrawal: excluding AI risk from coverage rather than setting the affirmative standards under which AI deployment could be insured, priced, and disciplined.

The issue is structural, not merely cyclical. Under the Berliner framework—the nine-criteria test the insurance industry has used to assess the insurability of emerging risks for more than four decades—generative AI currently fails outright on three criteria and is under strain on five more, according to a 2025 Geneva Association **assessment**. The most fundamental failure is information asymmetry. Carriers cannot see which models their insureds are running and for what applications, how those systems are governed, or whether claimed risk controls actually exist, and they therefore cannot price the two classical pathologies of insurance markets: moral hazard and adverse selection. Every other failure becomes easier to repair once that one is fixed—and none of them will fix themselves quickly. The cyber insurance market took roughly 20 years to mature, and it matured without ever developing the verification infrastructure that would have let underwriters discipline risk rather than merely price it. AI deployment is moving on a much faster clock, and the absence of insurance might pose a much greater adoption barrier.

CSIS | CENTER FOR STRATEGIC & INTERNATIONAL STUDIES

This paper proposes four recommendations, each tied to an existing institutional template and each aimed at a specific insurability failure. First, the federal government should build a two-track AI incident database at the National Institute of Standards and Technology (NIST), modeled on aviation's Aviation Safety Reporting System (this is the cheapest of the four recommendations and the one on which the other three depend because it repairs the missing loss-frequency and loss-severity data that all pricing requires). Second, the National Association of Insurance Commissioners should convene a federal-state working group before the current wave of AI exclusion templates hardens into a second generation of forms. Third, Congress should prepare—but stage—a layered, Price-Anderson-style federal backstop for catastrophic, correlated AI losses that the private market cannot absorb. Fourth, governments should license independent verification organizations (IVOs) and accelerate the auditable risk-assessment standards they verify against—the pairing that directly attacks information asymmetry and converts AI underwriting from guesswork into engineering.

None of these steps requires importing the European compliance model or the Chinese state-carrier model. Together, they would convert insurance from an accidental brake on American AI adoption into what insurance has historically been for every transformative technology since steam: an accelerator, and one of the most effective safety regulators the United States has.

Introduction

The 2026 closure of the Strait of Hormuz as part of the U.S.-Iran war is a straightforward case of Iran using military force—missiles, drones, and sea mines—to prohibit oil tankers and other sea vessels from safely transiting the geographic choke point to their intended destination port. More than a decade earlier, in 2012, however, the United States and its European allies were able to prohibit most of the world's tanker ships from carrying Iranian crude oil using a **different kind of weapon** aimed at a different choke point: financial sanctions to cut off access to maritime protection and indemnity (P&I) insurance. The insurance—which ensures that victims of liabilities such as spills, collisions, port damage, and injury will be properly financially compensated—is **mandatory** in order to be allowed to dock at nearly all major oil shipping facilities in the world. Without access to insurance, oil tankers carrying Iranian oil could not unload their oil cargos and were thus unwilling to take new ones. Iranian oil exports and associated government revenues accordingly **plummeted**.

The Iranian example vividly demonstrates how the presence or absence of insurance plays a critical role in expanding or constraining economic activity. Indeed, even for the 2026 conflict, insurance withdrawal was a major factor: P&I war-risk cover was **removed** for the strait on March 5, making the economic risk too high for shipowners to use it, even if they could find sailors willing to risk their lives. Companies both big and small depend on insurance as a vital prerequisite for engaging in an extraordinarily diverse set of business activities. Even outside the oil and shipping industries, it is utterly routine for business customers to refuse to work with suppliers who cannot prove they have the relevant insurance coverage. It is no exaggeration to say that, for many, perhaps most, businesses, activities that cannot be insured are effectively prohibited.

Today, amid the AI revolution, insurance is rarely at the top of mind of AI policymakers, but it belongs there. The presence or absence of insurance will be a major factor in accelerating or slowing U.S. adoption of AI technology. Moreover, the insurance industry frequently serves as a kind of private sector de facto regulator. For example, in the United States, fire and building safety codes are drafted

not by government agencies but by private standards bodies—chiefly the National Fire Protection Association, founded in 1896 by **fire insurance underwriters**, and the International Code Council. These model codes are the most powerful raw inputs used by legislators even though model codes carry **no legal force** until state and local governments adopt them by ordinance. State and local governments almost always do, however, because insurers reinforce them by grading and pricing how rigorously each community enforces them. Businesses and governments alike are strongly incentivized to adopt the insurance industry's preferred regulatory standards.

When it works well, this interplay between businesses, their insurers, and technology can lead to a virtuous circle with three simultaneous benefits: (1) a reduction in the risk of harms by incentivizing companies to adopt effective governance practices and risk mitigation technologies in pursuit of lower insurance premiums; (2) a reduction in the cost of mitigating potential harms by turning those governance practices and risk mitigation technologies into well-understood industry standards and best practices, and (3) accelerated overall adoption as more firms are willing and able to use the safer, cheaper, and better understood technology.

This is precisely why it is so critical that AI policymakers pay attention to what is currently going on in the insurance market. While most enterprise liability policies currently are assumed to have blanket coverage that includes activities involving AI, insurers appear poised to exclude some or all AI uses from the next round of renewals, viewing AI risks as too broad, diverse, and obscure to price correctly. If this happens, AI adoption in the United States will undoubtedly slow as companies find AI adoption uninsurable and thus unattractive.

In this regard, AI insurance policy is a form of AI industrial policy, and one that the United States needs to take seriously if it wants to maintain global leadership in the AI race. Indeed, China quite explicitly views insurance policy as a tool of industrial policy. China's government—firmly focused on accelerating AI adoption across the Chinese economy—just **announced** a host of measures to strengthen the availability, affordability, and effectiveness of enterprise AI insurance.

Section 1: The Regulator No One Elected

State insurance commissioner dockets do not normally make front-page news in technology outlets, but on April 23, 2026, that is exactly what happened, and with good reason. The Information **reported** that state insurance commissioners—the state government regulators that oversee and enforce insurance laws across the United States—had quietly approved more than 80 percent of requests from subsidiaries of large insurers to exclude AI-related damages from corporate insurance policies. In the case of AI, that currently is not much of a problem. The sorts of blanket insurance policies routinely purchased by businesses today—such as for product liability, professional liability, cyber insurance, and intellectual property (IP)—were for the most part written and issued before generative AI became the most **rapidly adopted** technology in human history. The blanket coverage statements of today's policies thus are **generally agreed** to cover AI.

That appears quite likely to change, and soon. The insurance companies' pursuit of prior regulatory approval for new types of policy exemptions very likely signals that many of the United States' largest insurance providers are securing the option to remove AI-related activities from their standard corporate insurance policies, much as they did with **cybersecurity-related incidents** more than a decade ago.

In a worst-case scenario, an un-insurability crisis could bring the unfolding AI revolution to a screeching halt. Adopting AI for business enterprise activities becomes far less attractive, and perhaps untenable, when the activities not only bring the risk of huge liability lawsuits but risks that cannot be insured against or may even void existing insurance policies.

For AI activities to be insurable, though, the insurance industry must have a credible methodology for quantifying the risks posed by different types of customers, different types of AI, and different types of use cases. At present, that is exceptionally difficult in the absence of widely accepted industry best practices. Imagine for a moment, that there was no basic understanding of which types of building materials (e.g., wood, brick, and concrete) were more or less likely to burn down. Would a company still be interested in selling fire insurance? Of course not.

> *If companies are prohibited from getting insurance for AI-related business activities, many of them will view that as nearly equivalent to such activities being prohibited by law.*

There is a bit of dramatic irony in the whole situation. The U.S. government has been pursuing a "light-touch" regulatory approach, arguing that the private sector AI industry should not be unduly burdened by government regulations and governance standards. However, the absence of broadly accepted AI governance standards—the sort of thing that allows a company to credibly say that it is adequately managing risk and an insurance company to robustly assess such claims—now risks putting the insurance industry in the position of a sort of private sector regulator without the tools to perform such a function. If the private sector insurance industry collectively decides to provide little to no AI insurance coverage to businesses, the de facto outcome would—in a worst-case scenario—nearly resemble that of strict government regulation. If companies are prohibited from getting insurance for AI-related business activities, many of them will view that as nearly equivalent to such activities being prohibited by law.

For two years, the visible U.S. conversation about AI regulation has unfolded in two arenas. The first is the international AI safety summit circuit. The second is the patchwork of state legislative fights, such as Colorado's SB 24-205 and New York's RAISE Act. Executive actions such as the Biden administration's October 2023 executive order, its rescission by President Trump in January 2025, the subsequent AI Action Plan, and the Trump administration's use of **export controls** to restrict Anthropic's Mythos 5 and Fable 5 models have absorbed most of the rest of the policy oxygen.

Unless major legislation passes, very few of the rules that will determine what most American businesses can actually do with AI in 2027 are being written in any of these places. They are being written in insurance commissioner filings across 50 states, integrated into Insurance Services Office (ISO) endorsement language, and accounted for in reinsurance treaty wordings. They are being written by carriers and reinsurers responding rationally to litigation expected liability curves their actuaries cannot yet price. And they are being written, by default, into a posture that constrains AI deployment rather than disciplines it.

This is not an indictment of the carriers. They are reading the same data as everyone else, and the data is troubling. Cumulative U.S. lawsuits related to generative AI grew 978 percent between 2021 and 2025, with year-on-year growth of 137 percent in 2024–25, according to litigation **data** compiled by Testudo Global. OpenAI is being **sued** by the *New York Times* and a coalition of authors for copyright infringement, as well as **by the parents** of a 16-year-old who died by suicide after extended conversations with ChatGPT. In June 2026, a coalition of state attorneys general **opened an investigation** into OpenAI related to a wide-ranging set of company activities.

This does not even get into the risks faced by industries such as finance, health, manufacturing, and professional services, in **all of which** companies are facing potential liability lawsuits related to their own use of generative AI.

But the cumulative effect of these rational individual decisions is a de facto regulatory regime, and a harsh one at that. In the United States, insurance has long functioned as the load-bearing regulator of risk-creating industries that statutory regulation either could not reach or with which it could not keep current. Fire safety code is enforced principally because building owners **cannot obtain** fire insurance without compliance. Workers' compensation, automobile liability, environmental liability, marine cargo, and aviation **each owe** more of their day-to-day operational discipline to the conditions insurers attach to coverage than to any black-letter legal statute. As Kevin Kalinich, Aon's intangible-assets global collaboration leader, **told** the Senate Banking Subcommittee on Securities, Insurance, and Investment on July 30, 2025: "Insurance does more than transfer risk; it incentivizes behavioral changes, imposes performance standards, and validates operational discipline. In fact, underwriting standards often act as a form of guardrails, setting the threshold for what is considered an acceptable risk."

What is unusual about the current AI moment is not that insurance has the potential to act as a de facto regulator. That is the **historical pattern**. What is unusual is that, for most of the market, it is regulating by withdrawal rather than by affirmative standard setting. A small specialty market among insurers is beginning to offer the affirmative bargain: CFC embedded **explicit AI coverage** across seven product lines in June 2026, Munich Re insures the performance of an **AI mortgage-screening tool** that it has itself evaluated, and Beazley underwriters now ask **structured AI-governance questions** as a condition of coverage. But these products remain boutique, with limits that top out at **$25–$50 million** per insured against litigation exposures that already run into the billions. The standardized policy forms that govern most enterprise coverage are moving in the opposite direction. For example, more than 60 property and casualty insurance providers in 2026 **filed** for AI exclusions in their upcoming policies. There, carriers are not telling enterprise customers, "we will cover your AI deployment if you adopt these governance practices and pay a calculated premium." A growing number are securing the right to say "**we will not cover your AI deployment at any price**" or "**we will only provide coverage that is a fraction of your total needs**." A regulatory posture of that kind has consequences for industrial policy. The United States is currently absorbing those consequences without having seriously debated them, and without having decided whether they are the consequences it wants.

This paper argues that insurance has become the de facto most important regulator of U.S. AI deployment, that the current trajectory of that regulation is producing an outcome neither carriers nor policymakers nor AI developers would have chosen deliberately, and that federal policy has a small

number of well-defined recommendations available to convert insurance from a brake on AI into a virtuous-cycle accelerator. What follows are six key judgments:

1. **Generative AI fails three of the nine criteria of the Berliner insurability framework outright**—maximum possible loss, loss frequency, and information asymmetry—and is at risk on the other six. Traditional risk-transfer structures cannot price it confidently.
2. **Pre-AI-priced legacy policies are silently exposed to AI losses.** State insurance commissioners have approved more than **80 percent** of carrier requests to strip AI from those policies. The next renewal cycle could be the first in which AI exclusions are the norm rather than the exception.
3. **The cyber insurance trajectory is the dominant industry mental model and precedent for what comes next in AI.** The early specialty AI carriers—CFC, Munich Re's aiSure, Armilla AI, Testudo Global, and Artificial Intelligence Underwriting Company—are following the cyber playbook closely, mirroring the example of the narrow exclusions in the late 1990s and standalone specialty market in the 2010s.
4. **Frontier AI labs cannot fully insure their own litigation exposure.** OpenAI is reportedly capped at roughly $300 million in coverage; Anthropic has agreed to pay a **$1.5 billion settlement** from its balance sheet and is reportedly considering tapping investor funds for future settlements. Vendor liability caps push residual risk down to enterprise deployers, and from deployers to consumers.
5. **Foundation-model concentration creates a systemic accumulation risk that no insurer has yet priced credibly.** A single model failure could trigger thousands of correlated losses across unrelated policyholders—a structure that defeats traditional reinsurance pooling and resembles terrorism or pandemic risk more than ordinary casualty exposure.
6. **China is using state insurance policy to accelerate AI adoption, while the United States is unintentionally constraining it.** Federal policy can convert insurance from a brake to an accelerator, but only with deliberate intervention along several specific lines—a federal qualification framework in the style of the SAFETY Act, a TRIA-style backstop for catastrophic systemic AI losses, federal coordination on AI-incident data, and a small number of additional recommendations laid out in Section 11.

Section 2: Insurance as Regulator: The Historical Pattern

Insurance has been the load-bearing regulator of risk-creating industries in the United States for more than a century. The pattern is consistent enough to constitute a structural feature of American political economy: When a new technology, business practice, or industrial process generates third-party harms faster than legislatures can write rules, the work of disciplining those harms tends to migrate to the insurance underwriting desk.

Massachusetts **mandated** automobile liability insurance in 1927, the first state to do so; every other state except New Hampshire eventually followed. Auto insurers funded the **Insurance Institute for Highway Safety**, founded in 1959, whose crash-test ratings have shaped vehicle safety design for decades.

The mechanism is the same in each cycle. Insurers operate under capital-adequacy requirements; they must hold financial reserves against potential losses, and they price coverage by aggregating actuarial data across their book of business. When a class of risk becomes uneconomic to underwrite at the existing terms, carriers either narrow coverage, raise premiums, or impose conditions on the insured. Insureds, in turn, adopt the conditions because the alternative is operating uninsured, which is rarely viable for a modern enterprise. The conditions—including loss-control programs, safety committees, vendor due-diligence protocols, and governance documentation—become, in practice, the standards the industry follows.

This raises a legitimate concern. There is something uncomfortable about delegating public safety standard setting to private actors who are neither democratically accountable nor required to weigh public interests beyond their own loss ratios. State insurance regulators monitor solvency and consumer protection; they do not generally direct what risks the industry chooses to underwrite. The result is a regulatory architecture whose contours reflect the underwriting appetites of a small number of capital-pool managers rather than a deliberative public process.

The rebuttal is that the alternative—comprehensive federal regulation written in statute—has been frequently either politically infeasible or technologically obsolete by the time it was enacted. Cyber is the clearest example. Federal cybersecurity legislation was debated continuously from the mid-1990s through the 2020s; the major statutory instruments that did pass—the **HIPAA Security Rule** (2003), the **Gramm-Leach-Bliley Safeguards Rule** (2003), and the **Cyber Incident Reporting for Critical Infrastructure Act of 2022** (CIRCIA)—addressed narrow sectoral slices. Through that period, what corporate cybersecurity actually looked like in practice was determined by what cyber underwriters required of insureds. Insurance was not a second-best regulator. It was the only regulator that operated at the speed of the technology.

For AI, the existing affirmative regulatory architecture is similarly thin. The most developed instrument is the National Association of Insurance Commissioners' **Model Bulletin** on the Use of Artificial Intelligence Systems by Insurers, adopted at the NAIC's Fall National Meeting on December 4, 2023. The bulletin sets governance, testing, and bias-monitoring expectations—but it applies to insurers' own use of AI in their own operations, not to the AI applications that other industries are deploying and that insurers are now declining to cover. Beyond the NAIC bulletin, there is no federal framework, no national insurer standard for evaluating insureds' AI deployments even on an industry-by-industry basis, and no analogue to the cyber underwriting questionnaires (and later third-party assessments) that became the de facto IT-security checklist in the 2010s.

What is unusual about AI is therefore not that insurance is acting as a regulator. It is that insurance is acting as a regulator without an affirmative product. In every prior cycle, workers' compensation, automobiles, environmental, and cyber, insurers eventually built coverage forms that imposed governance conditions in exchange for capacity. The market reached an equilibrium in which insureds could obtain coverage if they accepted the underwriting terms—usually corporate governance restrictions involving adhering to some management process implementation of some safety technology. The current AI cycle has not reached that equilibrium. The first move (exclusion) has happened. The second move—affirmative *standardized* coverage with governance conditions—has not, though the early specialty products profiled in Section 8 are the first attempts.

> *Any policy that accelerates the development of robust governance standards is one that will accelerate the adoption of AI as a whole.*

Kevin Kalinich, intangible-assets global collaboration leader at the insurance broker Aon, **told** the Senate Banking Subcommittee on July 30, 2025, "where AI governance is appropriately robust, growth and adoption are stronger and faster—not weaker and slower." What is needed, therefore, is a set of robust governance standards and **evaluation infrastructure** that can authoritatively and accurately clarify whether or not enterprise risks associated with AI are being appropriately identified, governed, and, where possible, mitigated. Any policy that accelerates the development of robust governance standards is one that will accelerate the adoption of AI as a whole.

Section 3: The Diagnosis: Why AI Fails the Insurability Test

The virtuous cycle described earlier in this paper (governance practices producing lower premiums and therefore producing fewer losses and faster adoption) requires that there be a coverage product on offer in the first place. For most categories of AI risk in 2026, there is not. Some companies, such as **Armilla**, are trying to remedy this with boutique products, but these remain small and outside the purview of the large, standardized policies. The reason is partly structural. AI currently does poorly on the basic test that the insurance industry uses to determine whether a category of risk can be insured at all.

That standard tool is the Berliner framework, named for the Swiss Re actuary Baruch Berliner, whose 1982 monograph *Limits of Insurability of Risks* set out nine criteria for assessing whether a class of risk is amenable to private insurance markets. The criteria fall in three groups, listed below:

Actuarial criteria ask whether the risk behaves in ways that allow underwriters to price it, including

- randomness of loss occurrence;
- maximum possible loss;
- average loss amount;
- loss frequency; and
- information asymmetry.

Market criteria ask whether the market for that risk can clear at sustainable terms, including

- adequate premiums; and
- acceptable coverage limits.

Societal criteria ask whether insurance for the risk is socially and legally tenable, including

- consistency with public policy; and
- legal permissibility.

In October 2025, the Geneva Association—the international think tank for the insurance industry—applied the Berliner framework to generative AI risks across all nine criteria, drawing on a **survey** of 600 corporate insurance decisionmakers across the six largest insurance markets: China, France, Germany, Japan, the United Kingdom, and the United States. Three criteria received an outright red

verdict, indicating violation of traditional insurability under existing risk-transfer structures. Five received yellow, indicating significant challenges short of outright violation. Only one received green. The full assessment is reproduced below in Table 1.

Table 1: Berliner Framework Assessment of Generative AI Insurability

Category	Criterion	Assessment	Verdict
Actuarial	Randomness of loss occurrence	Generative AI introduces new layers of complexity to avoid producing hallucinations or harmful content, which makes it difficult to assess whether failures occur randomly when systematic validation is lacking.	Yellow
Actuarial	Maximum possible loss	Wrong or malicious code generated by AI can lead to massive service disruption, potentially causing systemic risk. Generative AI failures such as spreading misinformation, IP violations, and deepfake-driven fraud in critical sectors (e.g., healthcare and finance) can also lead to large losses, particularly when the failure persists for a long time or is subject to regulatory penalties.	Red
Actuarial	Average loss amount	Generative AI incidents can lead to high potential financial and reputational damage, such as from misinformation or regulatory fines.	Yellow
Actuarial	Loss frequency	Context-dependent risks with limited data make frequency estimation difficult, hindering diversification. In the medium term, sufficient frequency can be expected.	Green
Actuarial	Information asymmetry	Insured parties may neglect AI system integrity (moral hazard), while riskier AI systems may seek coverage (adverse selection). Insurers may struggle to verify generative AI risks and how businesses manage them.	Red
Market	Adequate premiums	Heightened uncertainty about processes underlying generative AI and the associated ways in which harm can arise may imply additional premium loading and impose pressures on affordability for smaller businesses relying on these technologies.	Yellow
Market	Acceptable coverage limits	Insurers may hesitate to offer high limits for generative AI due to uncertainty in potential liabilities (widespread content misuse or errors in AI-generated decisions).	Yellow
Societal	Consistency with public policy	Generative AI raises ethical issues, such as creating harmful or biased content, which may conflict with societal norms and reduce policymakers' acceptance of certain insurance products.	Yellow
Societal	Legal permissibility	The evolving landscape of copyright, intellectual property, liability laws, and AI regulations makes it challenging to define insurable risks and underwrite them.	Yellow

Source: Geneva Association, *Gen AI Risks for Businesses: Exploring the role for insurance* (Geneva: Geneva Association, October 2025), Table 2, p. 28, https://www.genevaassociation.org/sites/default/files/2025-10/gen_ai_report_0110.pdf. Reproduced with permission of the Geneva Association.

Three of the red verdicts deserve elaboration in prose because each one names a distinct mechanism by which traditional risk-transfer structures break down for AI.

The first is **maximum possible loss**—the largest plausible payout against a single triggering event. Conventional insurance pools risk on the assumption that losses are bounded and approximately independent across insureds. AI fails both legs. A single defect in a foundation model could propagate simultaneously across every enterprise running that model, whether for a hallucinated medical diagnosis pattern, a mishandled regulatory disclosure, or a bias in a hiring algorithm. When a single failure can cascade across 1,000 or 10,000 correlated policyholders, the pooling of uncorrelated risk logic that makes private insurance possible begins to fail.

The second is **average loss amount**—the expected severity of the typical claim, and the number from which premiums are actually priced. Mature lines of insurance rest on tight, well-documented severity distributions: Decades of claims data tell an underwriter what a warehouse fire or a rear-end collision usually costs. Generative AI has no such distribution. The same underlying failure mode—a system confidently generating false content—has produced losses of **CAD 812** for Air Canada's chatbot, **$25 million** for the engineering firm Arup after an AI-generated deepfake of its executives, a **$110 million** defamation claim against Google over its AI Overviews feature, and a **$1.5 billion** copyright settlement for Anthropic. That is a severity range spanning six orders of magnitude, before adding regulatory penalties that the EU AI Act sets as high as **7 percent** of global annual turnover. A line of business with no stable average loss is a line of business where the premium is a guess.

The third is **information asymmetry**. The insured knows more about its own AI deployment—such as the model in use, the training data, the prompt engineering practices, and the human-in-the-loop discipline—than the carrier ever can. That asymmetry creates the two classical pathologies underwriters worry about. The first is moral hazard: An insured may neglect AI system integrity once it is covered. The second is adverse selection: Insureds with the riskiest AI deployments are the ones most willing to pay for coverage. Neither pathology is new. They are the canonical failure modes of insurance markets that George Akerlof, Michael Rothschild, and Joseph Stiglitz **formalized in the 1970s**—work that earned the 2001 Nobel Prize in economics. The theory's core prediction is the one now playing out in AI: When the sellers of coverage cannot verify what the buyers know, coverage shrinks or disappears entirely. Underwriters cannot currently independently verify how an insured is using AI, and they cannot price these pathologies out without a verification regime that does not yet exist. The Lloyd's Market Association's **mid-2025 survey** of 144 market participants confirmed the diagnosis from inside the market: "The survey responses are informed by views and opinions rather than data in most cases" because "insurers are remote to insureds' use of AI; it is not always clear whether an insured is using AI in the ordinary course of their business and little data on AI usage is currently collected during underwriting."

The remaining five yellow criteria are not free passes. They flag affordability pressures on smaller businesses, hesitation about high coverage limits, ethical concerns over harmful or biased content, and a regulatory environment whose contours are still moving.

> *The industry is not refusing to engage with AI because it does not see the demand. It is refusing to engage because the underwriting science required to meet the demand at sustainable terms does not yet exist.*

This diagnosis matters because the demand for AI insurance is real, and the availability or absence of AI has major implications for AI adoption. The Geneva Association survey found that more than 90 percent of corporate insurance decisionmakers across the six largest insurance markets express a need for coverage tailored to AI threats, and more than two-thirds would pay at least 10 percent more in premiums for explicit generative AI policy extensions. At Lloyd's, managing agent boards **now rate** AI risk as second only to geopolitical risk on their risk registers, and 86 percent of underwriters expect AI usage among insureds to increase over the next two to three years. The industry is not refusing to engage with AI because it does not see the demand. It is refusing to engage because the underwriting science required to meet the demand at sustainable terms does not yet exist.

That diagnosis is the analytic foundation for everything that follows. Section 4 examines how carriers are responding in the short term—by stripping AI from existing policies through endorsement-driven exclusions and sub-limit caps. Section 8 traces the longer-term path the industry is now committed to, modeled on the cyber insurance precedent. The policy implications are laid out in Section 11. But none of those moves change the underlying problem: Under the framework the industry has used to test the insurability of every other emerging risk for 44 years, AI does not yet pass.

Section 4: The Retreat: How Carriers Are Stripping AI from Existing Policies

The clearest illustration of what the diagnosis in Section 3 means for carrier behavior comes from a single sentence in a W.R. Berkley regulatory filing first **reported** by the *Financial Times* in November 2025. The carrier's proposed endorsement bars coverage for any claim involving "any actual or alleged use" of AI, including any product or service "incorporating" the technology. The language of "any actual or alleged" is extraordinarily broad: A claim does not have to involve actual AI use; it merely has to involve alleged AI use. Under that wording, a plaintiff need only assert AI involvement, however weakly, for the carrier to invoke the exclusion. The exclusion is doing the underwriting work that an underwriting model cannot.

Lockton Re **stated** the same finding in plainer language in February 2026: "It is clear that CGL insurers do not currently model, underwrite, or price AI risks, so there is likely a growing gap between what insurers intend to cover and what they actually cover based on the policy language."

The retreat is taking two forms in parallel. The first is endorsement-driven exclusion: Carriers file new endorsement language with state insurance commissioners that strips AI exposures from existing commercial general liability (CGL), errors and omissions (E&O), and directors and officers (D&O) policies. The second is sub-limit caps: AI-related losses remain nominally covered under existing cyber and E&O policies, but only up to a small fraction of the policy's overall limit. Each mechanism is rational on its own actuarial logic. Taken together, they produce a coverage landscape in which the

gap between what coverage businesses think they have and what they actually have is widening with every renewal cycle.

THE ENDORSEMENT FILINGS

The endorsement story can be pinned to specific dates and specific carriers. The Insurance Services Office (ISO), the Verisk subsidiary that publishes standardized policy language used by U.S. admitted-market carriers, introduced generative AI exclusion endorsement templates in 2025; the ISO Generative AI Exclusion for commercial general liability **took effect** on January 1, 2026. The carriers actively filing AI exclusions include Berkshire Hathaway, Chubb, Travelers, AIG, Tokio Marine Holdings, W.R. Berkley, Great American, and Fairfax Financial. AIG's filing characterized generative AI as a "wide-ranging technology" whose claim exposure will "likely increase over time"; the company has **stated** that it has no current plans to implement the exclusion but has secured the regulatory option to do so.

In its October 2025 report, the Geneva Association reproduced an analysis adapted from earlier broker work by Aon. The analysis maps 14 categories of AI risk against the eight principal lines of corporate insurance. The matrix is the visualization of the retreat. Of 112 cells in the table, only 13 record an "Available" verdict and another 9 are "Limited." The remaining 90 cells, roughly 80 percent of the matrix, are "Excluded." The full table is reproduced with the permission of the Geneva Association below.

The pattern is consistent with the broader story. Damage claims relating to bodily injury and tangible property damage arising from AI defects (at least for now) remain available under product liability and CGL. Privacy and security breaches involving AI remain available under standalone cyber. Employment discrimination claims involving AI hiring tools remain available under employment practices liability. Beyond those handful of cells, the affirmative coverage is sparse. Patent infringement is covered only under specialty IP policies. Some AI perils are excluded everywhere. Loss of financial assets requires a separate crime policy that most enterprises do not carry. The matrix is, almost everywhere it is read, a map of what is not covered.

The carriers are not retreating at random. They are responding to a litigation curve that has run faster than any actuarial model built before 2024 anticipated. **Litigation data** compiled by Testudo Global Inc. and reported by Gallagher Re shows that cumulative U.S. lawsuits involving generative AI grew 978 percent between 2021 and 2025, with year-on-year growth of 137 percent in 2024–25. Wolfe Research separately **estimates** that approximately 800 consumer AI-related lawsuits were filed against U.S. businesses in 2025, up 140 percent year-on-year. The Stanford AI Index **reports** 362 tracked AI-related incidents in 2025, up 55 percent since 2024.

The next section takes up the prior question: Who, if anyone, is supplying coverage to the frontier AI labs themselves, whose own litigation exposure already exceeds the capacity available in any current insurance market?

Table 2: AI Insurance Coverage Gaps Across Principal Lines of Corporate Insurance

AI peril	Media liability	Tech E&O / PI	Product liability	General liability	IP	Standalone cyber	D&O	Employment
Third-party damages liability for faulty products or services	**Excluded**	**Limited**	**Available**	**Limited**	**Excluded**	**Excluded**	**Excluded**	**Excluded**
Copyright, trademark, or service-mark infringement	**Available**	**Limited**	**Excluded**	**Limited**	**Available**	**Excluded**	**Excluded**	**Excluded**
Patent infringement	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Available**	**Excluded**	**Excluded**	**Excluded**
Discrimination	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Available**
Defamation, libel, slander	**Available**	**Limited**	**Excluded**	**Limited**	**Excluded**	**Excluded**	**Excluded**	**Excluded**
Bodily injury	**Excluded**	**Excluded**	**Available**	**Available**	**Excluded**	**Excluded**	**Excluded**	**Excluded**
Tangible property damage	**Excluded**	**Excluded**	**Available**	**Available**	**Excluded**	**Excluded**	**Excluded**	**Excluded**
Privacy and security breaches	**Excluded**	**Limited**	**Excluded**	**Limited**	**Excluded**	**Available**	**Excluded**	**Excluded**
Loss of financial assets (requires crime policy)	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**
Market manipulation	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**
Autonomous weapon	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**
Product recall	**Excluded**	**Excluded**	**Limited**	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**
Business interruption	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Available**	**Excluded**	**Excluded**
Breach of directors' or officers' duties	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Excluded**	**Available**	**Excluded**

Source: Geneva Association, *Gen AI Risks for Businesses*, Table 4, p. 34. Reproduced with permission of the Geneva Association.

Section 5: The Labs Themselves: Capacity Short of Exposure

The retreat documented in Section 4 pushes insurance coverage off enterprise balance sheets. Section 5 takes up the harder problem at the top of the value chain. The frontier AI labs—the small set of companies whose foundation models the rest of the economy is now building on—face a structural mismatch between their litigation exposure and the insurance capacity available to absorb it. The mismatch is not marginal. It is roughly an order of magnitude.

The most concrete data point comes from a *Financial Times* **report** on October 8, 2025, drawing on people familiar with the policy: OpenAI has secured cover of "up to $300 million" for emerging AI risks, brokered through Aon. The cap is small relative to the lawsuits already in flight.

What the *Financial Times* reporting calls "multibillion-dollar claims" is therefore a literal description, not a figure of speech. Aon's Kevin Kalinich **said**: "We don't yet have enough [insurance industry] capacity for [frontier AI model] providers." Capacity in insurance language is not a euphemism for caution. It is the actual financial capital the global insurance and reinsurance markets are willing to commit to a single line of risk. For categories where capacity is mature—such as property, marine cargo, and large-account workers' compensation—a single Fortune 500 company can buy coverage in the multibillions. For frontier AI labs, the visible ceiling is in the low hundreds of millions. This gap reveals the current structural problem.

Both OpenAI and Anthropic have **reportedly explored** a fallback that has been a standard option for cyber and social media exposure for two decades: the captive insurance vehicle. A captive is a wholly owned subsidiary that an enterprise capitalizes and uses to insure its own risk, often passing part of that risk on to the conventional reinsurance market. Microsoft, Meta, and Google have used captives to manage some aspects of cyber and platform-liability exposures since the 2010s. The captive route makes sense on the balance sheet for individual companies—it allows a firm to self-insure on commercial terms while preserving tax and regulatory benefits—but it does not deepen the broader market. A captive is not a coverage product; it is a balance sheet partition. If every frontier lab self-insures through a captive, the insurance market remains as thin as it is today, and the deployer ecosystem downstream of the labs is no closer to coverage than it was before.

The structural consequence sits in vendor contracts. OpenAI's **standard business terms** cap its liability to an enterprise customer by writing "total liability under the agreement will not exceed the total amount customer paid to OpenAI during the twelve months immediately prior to the event giving rise to liability." Anthropic and other foundation model providers **offer** comparable terms. Gallagher Re **notes** that under this arrangement "AI vendor contracts favor the vendor and leave deployers shouldering most of the liability," typically capping liability at 12 months of fees with no performance warranties. There is a **noteworthy exception** for IP violation liability, in which OpenAI, Anthropic, Microsoft, and Google Vertex AI indemnify their customers against claims. However, this vendor indemnity does not cover outcomes such as wrongful death, bodily injury, employment discrimination, defamation, market manipulation, AI-driven fraud, AI-driven cyberattacks, regulatory penalties, and autonomous action torts. Those are the categories where the broad 12-month-of-fees liability cap still applies. This clause does what it is designed to do: It allocates residual risk to the customer of the frontier model providers. When the customer is a Fortune 500 deployer, the deployer either absorbs the cost or pushes it on to its own customers. Even at that scale, however, the negotiating leverage

is thinner than the balance sheet suggests. With only a handful of frontier model providers to buy from, large deployers routinely report accepting liability allocations they would refuse from any other category of vendor, as the practical alternative is not deploying frontier AI at all. When the customer is a small enterprise without bargaining power or a sophisticated risk management function, the residual risk lands on a balance sheet that cannot absorb it. ***Mobley v. Workday*** is the live test case for what happens when a deployer, not the model developer, becomes the named defendant in an algorithmic discrimination class action lawsuit.

There is a legitimate defense of how the labs are managing this. The risk is genuinely uncertain. The actuarial data does not yet exist. Frontier labs that cannot price their own exposure cannot be expected to assume their customers' exposure on top of it. Aggressive vendor terms are the rational response to a market in which the underwriting science has not caught up to the technology.

But the cumulative effect of these rational decisions is a regressive distribution of AI risk down the value chain. Capital-rich frontier labs hold the residual risk on their own balance sheets at the top, often through captives. Capital-rich enterprise deployers absorb the next layer in the middle, sometimes through specialty products (examined in Section 8) that cap at $25-$50 million per insured. The remainder lands on smaller deployers and on consumers, who carry no insurance and no contractual recourse against the model developers whose outputs caused the harm. The Ada Lovelace Institute documents this pattern in detail in its **December 2025 study** of UK liability allocation: "Most businesses are faced with standard contractual clauses that shift liability risks away from large AI model vendors," with the practical consequence that small and medium enterprises in the middle of the value chain carry liability burdens that neither the developer nor the largest downstream deployer is willing to accept.

The next section deals with the question of why the available capacity stops where it does.

Section 6: The Unmodeled Peril: Systemic Accumulation

Aon's Kevin Kalinich offered the most concise diagnosis of the capacity problem in an **interview** with the *Financial Times* in November 2025: "What they can't afford is if an AI provider makes a mistake that ends up as 1,000 or 10,000 losses—a systemic, correlated, aggregated risk." Each adjective in that sentence is doing work: systemic, because the loss originates in shared infrastructure rather than in any single insured's behavior; correlated, because the losses arrive at the same time; and aggregated, because they sum into a number the market has never before had to absorb in a single line of business.

That structure is an important reason the visible capacity ceiling for frontier labs sits in the low hundreds of millions rather than the multibillions. Conventional insurance pools risk on the assumption that losses are approximately independent across insureds. Foundation-model concentration breaks that assumption. The same generative AI model—such as GPT, Claude, Gemini, or Llama—is now embedded in hiring, lending, healthcare, legal research, customer service, and enterprise productivity software across thousands of unrelated companies. A single defect in one of those models—such as a hallucinated medical diagnosis pattern, a mishandled regulatory disclosure, or a bias in a hiring algorithm—might not merely produce a single loss at a single insured. It could produce correlated losses at every insured running the model. This is the pattern the **2024 CrowdStrike outage** demonstrated for conventional software.

Traditional reinsurance pooling, which works by diversifying loss across uncorrelated risks, does not function in that setting. The pool turns out to be one large insured rather than a thousand small ones.

The structural analogy in the insurance industry's own working vocabulary is not commercial-line casualty risk; it is terror risk, pandemic risk, and nuclear risk—the three categories in which the private market has historically required a federal backstop because correlated tail losses exceeded the capital the global reinsurance market was willing to commit.

The history is instructive. The Price-Anderson Nuclear Industries Indemnity Act of 1957 **created** a federal backstop for nuclear power plant operator liability after the private insurance market made clear it would not insure commercial nuclear power at any price the utilities could pay; the act has been renewed continuously since, most recently extended to 2065. Pool Re was created in the United Kingdom in 1993 after the Provisional Irish Republican Army bombings of the Baltic Exchange (April 1992) and Bishopsgate (April 1993) drove commercial property reinsurers out of the UK terrorism market and forced the British government to assemble a **mutual reinsurance pool** to keep central London insurable. The **Terrorism Risk Insurance Act of 2002**, signed in the wake of the September 11 attacks, established a federal reinsurance backstop for catastrophic terrorism losses in the United States. In each case the trigger was the same: a category of correlated tail risk too large for the private market to absorb on its own balance sheet, and an industry willing to write the line only if the federal government wrote the upper layer.

Yoshua Bengio, the deep-learning researcher who shared the 2018 ACM A.M. Turing Award with Geoffrey Hinton and Yann LeCun, has **called** for the United States and other governments to mandate liability insurance for AI companies modeled on the nuclear template. Bengio's reasoning is that frontier AI development presents a class of catastrophic risk exposure analogous to nuclear power generation, and that the discipline of compulsory liability insurance—and the underwriting due diligence that would come with it—would force frontier developers to internalize the safety investments their operations require.

The proposal deserves serious engagement, including by readers who do not share **Bengio's beliefs** about the catastrophic risks of AI. The nature of AI risk need not be catastrophic or existential for it to share structural features with the nuclear and terrorism examples. The structural fact that the underlying market is failing to clear at sustainable terms (established in Sections 3, 4, and 5) is independent of how one weights existential risk. But there is a meaningful analytical limitation to the nuclear analogy that Bengio's framing tends to elide.

The Price-Anderson regime is a narrow instrument. It covers a small and identifiable population, roughly a hundred U.S. commercial reactors. These are operated by a small number of regulated entities, monitored by a single federal agency (the Nuclear Regulatory Commission), and physically sited in known locations. A single accident at a single plant produces a single localized loss event. The federal backstop kicks in if and only if commercial nuclear plant operator liability exceeds the private retention layer, and the trigger is comparatively easy to certify because everyone knows where the reactors are.

> *A single model defect produces correlated losses across thousands of geographically distributed insureds with no shared regulator and no shared technical interface.*

Foundation model risk has none of those structural properties. The relevant "installations" are not power plants; they are millions of business-process integrations across every sector of the economy, deployed by tens of thousands of entities on infrastructure operated by a handful of model providers and a much larger number of cloud and middleware vendors. A single model defect produces correlated losses across thousands of geographically distributed insureds with no shared regulator and no shared technical interface. An important share of the AI risk is concentrated at the source—at the foundation model layer—but distributed at the loss event. That asymmetry matters for insurance instrument design: A Price-Anderson clone rebuilt for AI would need to look very different from its 1957 ancestor, with a new approach to the certification trigger and a new approach to allocating retention across the value chain.

What does transplant cleanly from the nuclear, terror, and pandemic precedents is the structural insight that some categories of correlated tail risk cannot be absorbed by the private market alone. Section 11 takes up the policy mechanism—a TRIA-style federal reinsurance backstop, scoped to certified catastrophic AI-loss events above an industry-wide retention—that follows from that diagnosis. Section 8 takes up first the question of what the private market is actually building, in the form of standalone AI insurance products, and whether the trajectory of that build-out is fast enough to matter without federal intervention.

Section 7: Autonomous AI Hacking and Its Insurance Implications

On July 16, 2026, Hugging Face **disclosed** that an intruder had broken into its production systems, stolen credentials, and moved through its internal clusters for the better part of four days. While cyber breaches at this point are somewhat routine, this one was different. The reason took another five days to surface. On July 21, OpenAI **acknowledged** that the intruder was one of its own AI models.

The models were being tested for cyber capability in a sealed environment, with their safety guardrails switched off for the exercise. They were supposed to solve a hacking benchmark. Instead, they did what a student who is short on time sometimes does. They went looking for the answer key. The models found a previously unknown flaw in the proxy meant to contain them, used it to reach the open internet, and broke into Hugging Face, where they **believed** the test answers were stored. No customer data was touched and nothing was stolen. Hugging Face chief executive Clément Delangue **said** that it was "mind-blowing that all of this happened autonomously."

This was not a one-off. In early August, Meta **disclosed** that one of its models had reached the open internet and hacked an outside company during a similar test. On August 4, the UK AI Security Institute **reported** that OpenAI and Anthropic models had taken "autonomous, unsanctioned action on the live internet" in 10 of 122 safety-test runs. In one of those runs, a model fabricated identities to trick a human into approving a software update that concealed malware. Criminals are already copying the method. In July, the security firm Sysdig **documented** JADEPUFFER, the first ransomware operation

run start to finish by an autonomous agent. A human picked the target, and the AI agent executed the remainder of the steps autonomously. Anthropic had **flagged** the direction in November 2025, when it disclosed the first-known AI-orchestrated espionage campaign, run by a Chinese state actor. What was a single disclosure last fall is now a monthly event.

The offensive cyber capabilities of AI are getting stronger at the same time. Anthropic's Mythos model can **find software vulnerabilities** far faster than any human and has already surfaced thousands of previously unknown flaws in widely used operating systems and browsers. Anthropic judged that capability dangerous enough to lock it behind **Project Glasswing**, a vetted-access program launched in April 2026. In May, the International Monetary Fund **warned** that AI-driven cyberattacks had become a threat to financial stability. Simon Hughes, chief commercial officer at the cyber insurer Cowbell, put the underwriting problem in one line: "The scale at which [Mythos] can do that is beyond what anyone is capable of defending against."

Mythos creates a second problem for insurers, quieter than the first. Because access is restricted, underwriters cannot see the model well enough to judge what it can do. A June 2026 **Gallagher Re analysis** concluded that restricted release and saturated public benchmarks leave the market unable to evaluate frontier models at all, and that insurers are left to "**price uncertainty rather than risk**." The same analysis found that the time between a vulnerability's disclosure and its exploitation has fallen from more than two years in 2018 to roughly 10 hours in 2026. This is the information asymmetry problem from Section 3, only in a harder form.

Thus far, none of the summer's breakouts **caused substantial real-world harm**, and the cyber market has absorbed fast-moving threats before. But the incidents break an assumption implicitly included in nearly every cyber pricing model: that running a sophisticated cyberattack is expensive for the attacker. As CyberCube's William Altman **put it**, "ransomware used to require a team, and the cost of paying that team limited how many attacks were worth running and who was worth attacking." If advanced AI can reduce the cost of the attack team to nearly zero, then the small business that was never worth a criminal's time becomes a target like any other.

Thus far, the evidence of the insurance market reaction is once again exclusion. On July 10, 11 days ahead of OpenAI's disclosure, Verisk's Insurance Services Office **confirmed** it was drafting coverage exclusions for agentic AI. What coverage survives is somewhat murky. A **post-mortem** written with input from roughly 700 chief information security officers told them to go read how their cyber and technology E&O policies define the word "user." If it means an employee or a contractor, a rogue agent's actions may not be covered at all. A July report from the Artificial Intelligence Underwriting Company, **co-authored** with researchers from Anthropic and OpenAI, put more than 90 percent of insurers' exposure to AI agents in silent coverage that no one had priced. At the time, Hugging Face had no contract with OpenAI and thus little explicit leverage. It has **asked** OpenAI for a full log of the agents' actions and $100 million in compute for community defenses, and OpenAI has not agreed. "The liability is real, it is growing, and it sits on nobody's books," wrote New York University finance professor Haran Segram.

Section 8: The Cyber-Insurance Precedent: A Market in Formation

The dominant working analogy inside the insurance industry for what comes next in AI is not nuclear, terror, or pandemic. It is cyber. Gretchen Hoff Varner, the Covington and Burling attorney who has advised carriers on AI exclusion endorsements, **said** "As we started to see data breaches, computer-enabled losses, and subsequent class actions, insurers made a business decision to exclude cyber losses and create an entirely new category of product. We are seeing that same playbook applied to AI." The same framing appears in the Geneva Association's October 2025 report, in Lockton Re's February 2026 white paper, and across multiple *Financial Times* pieces in late 2025 and 2026. The exclusion comes first; the standalone product comes after.

THE CYBER SEQUENCE

The cyber timeline is recent enough to be checked against current carrier behavior. The first significant forms of cyber insurance policy appeared in the late 1990s, with what the Geneva Association in retrospect described as "narrow protections" and "very low coverage limits" written by underwriters "feeling their way through uncharted territory." Through the 2000s and 2010s, the ISO progressively introduced electronic-data and cyber-exclusion endorsements stripping digital perils from standard commercial general liability, and a specialty cyber market grew up behind the exclusions. After the Target, Sony Pictures, Anthem, and Equifax breaches between 2013 and 2017 gave the industry years of accumulated loss data to price against, standalone cyber transitioned to a recognizable mainline coverage with several billion dollars in annual U.S. gross written premiums. The exclusions and the new product co-evolved over roughly two decades. The policy form did not arrive until the actuarial science arrived to support it.

The cyber precedent is not, however, an unambiguous success story, and the strongest skeptical account deserves engagement in its strongest form. In a **2025 study** of what AI policymakers should learn from the cyber insurance record, Daniel Schwarcz of the University of Minnesota and Josephine Wolff of Tufts University conclude that cyber insurance largely failed in the regulatory role this paper describes: Carriers competed on price and coverage breadth rather than on security discipline, underwriting requirements remained shallow and inconsistently enforced, and insurers' routine reimbursement of ransom payments arguably financed the very ransomware economy they were underwriting against. Wolff's **book-length history** of the market reaches the same verdict at greater depth. On this account, cyber insurance "matured" as a commercial product without ever maturing as a private regulator.

That record strengthens rather than weakens the argument of this paper. What the skeptics document is a market that formed without the infrastructure that regulation-by-insurance requires: Cyber underwriters never acquired verification tools that could see inside their insureds' security postures, never pooled incident data systematically, and never converged on auditable standards—so they priced risk without disciplining it. Cyber insurance is thus less a reassuring precedent than a cautionary one: proof that the exclusion-then-product sequence can complete commercially while failing publicly. The recommendations in Section 11 are designed to build for AI, at the outset, precisely the standards, verification, and incident-data infrastructure whose absence hollowed out cyber insurance's regulatory promise. Notably, Schwarcz and Wolff arrive at the same destination from the opposite direction: Their

central recommendation is that regulators mandate standardized AI incident data collection rather than wait for insurers to generate it.

THE EARLY SPECIALTY MARKET

The early specialty AI carriers are now visible. Munich Re's aiSure, **available** in the market since 2018, uses an explicit model performance trigger: The underwriter and the insured agree on a target accuracy or error rate metric for a defined AI deployment, and the policy responds when actual model performance falls below that threshold. Armilla AI, a managing general agent backed by the Lloyd's syndicates Chaucer and Axis Capital, sells standalone coverage for AI underperformance, hallucinations, inaccuracies, and regulatory violations, with limits up to $25 million per insured. Testudo Global launched its AI insurance product in January 2026 in partnership with Gallagher Re and MIT, targeting middle-to-large enterprise generative AI deployers with a claims-made litigation trigger. Artificial Intelligence Underwriting Company (AIUC), based in San Francisco, writes coverage for up to $50 million per insured and counts ElevenLabs among its named policyholders. In China, the People's Insurance Company of China (PICC) has **piloted standalone generative AI insurance** for IP, portrait, and reputational infringement claims.

The endorsement track is moving in parallel. AXA XL's CyberRiskConnect generative AI endorsement, launched in October 2024, was the first significant AI-specific endorsement from a top-tier carrier. The endorsement covers data poisoning, usage rights infringement, and regulatory violations under the European Union's AI Act—but only for companies developing or deploying their own generative AI models, not for users of third-party services such as ChatGPT. The third-party-services exclusion is the analytical limit of the product. The customers most likely to need coverage are precisely the ones the endorsement does not cover. AXA XL's structure illustrates the broader pattern. Specialty AI products are easier to underwrite when the insured controls the model and the deployment because the underwriter can verify governance, training data provenance, and validation discipline. They are much harder to underwrite when the insured is a downstream consumer of a foundation model the insured did not build and cannot inspect.

CAPACITY AND THE VIRTUOUS CYCLE

Two structural disanalogies complicate the cyber precedent. The first is timing. The cyber market took roughly 20 years to develop from late 1990s niche policy forms to late 2010s mainline coverage, with actuarial science maturing only after the Target, Sony Pictures, Anthem, and Equifax breaches between 2013 and 2017 supplied the underwriting record. The AI specialty market does not have 20 years to wait. A specialty market that arrives at scale in 2046 is the wrong instrument for a deployment phase running through 2027.

The second is the structure of the correlated loss. Cyber loss accumulation is bounded by the IT architecture of individual organizations—a breach at one company does not automatically propagate to every other company running the same software stack. Foundation-model loss accumulation has no such bound. A single defect or attack vector in a widely deployed model could produce correlated losses across thousands of unrelated deployers, in the structural pattern documented in Section 6. The cyber market matured into a multibillion-dollar standalone line in part because reinsurance pooling worked at the loss-event layer; the AI market faces a tail-risk profile at the foundation-model layer that defeats the diversification logic on which reinsurance pooling is built.

That distinction matters for the question of whether the virtuous cycle Kalinich described—governance practices producing lower premiums and therefore producing fewer losses—will arrive on the timeline on which AI deployment is moving. The cycle requires not just that products exist, but that products are available to enough enterprises at limits high enough to make the underwriting conditions worth meeting, such that the conditions become a de facto industry standard. The cyber cycle, from the late 1990s emergence of niche policies to mainline standalone status by the late 2010s, took roughly 20 years. Section 9 takes up whether the European model offers an alternative.

Section 9: The European Union Is Also Failing

A once common objection to the diagnosis in the previous sections goes something like this: "The United States is letting insurance markets fill the void in AI liability law because Washington has elected not to legislate. Europe, by contrast, has the EU AI Act, the AI Liability Directive, and a regulatory tradition of comprehensive ex-ante rule-making that the United States lacks."

However, anyone arguing that the European Union has a solution for the AI liability issue is long out of date. The European AI regulatory framework was originally designed as two halves. The EU AI Act was the first half: ex-ante compliance rules to prevent harms. The **AI Liability Directive** was the second half: ex-post liability rules to compensate victims.

As of mid-2026, the first half has been delayed and diluted. The second half is dead.

Regarding the first half, the delay came on July 27, 2026, when the AI Omnibus—Regulation (EU) 2026/1744—entered into force, six days before the EU AI Act's high-risk rules were due to take effect. The omnibus pushed those rules back to December 2, 2027, for standalone high-risk systems and to August 2, 2028, for systems embedded in regulated products. The omnibus also eased compliance requirements for smaller companies.

Regarding the second half, the European Commission proposed the AI Liability Directive in September 2022. It **slated** the directive for withdrawal in February 2025, **formally withdrew** it on July 16, 2025, and **published** the withdrawal in the Official Journal of the European Union that October.

Still, the **EU AI Act** is a real law with real teeth. It entered into force on August 1, 2024. Its bans on prohibited AI practices took effect in February 2025, and its general-purpose AI rules took effect in August 2025. Fines reach €35 million or 7 percent of global annual turnover, whichever is higher. The omnibus did not repeal any of that. Transparency rules still **took effect** as scheduled on August 2, 2026, and the omnibus added new prohibitions on AI-generated child sexual abuse material and "nudifier" apps. No other major jurisdiction has enacted anything so ambitious.

The one liability instrument that did pass tells the same story. The European Union updated its **Product Liability Directive** (PLD) in 2024. The revised directive extends the definition of "product" to cover software and AI systems. That is a genuine doctrinal update. But as Andrea Bertolini, a professor of private law at Scuola Superiore Sant'Anna **puts it**, the reform "leaves intact the substantive limitations that rendered the PLD largely ineffective in the first place." The original 1985 directive produced few cases and fewer successful claims. Victims continue to mostly fall back on national tort and contract law. There is little reason to expect the 2024 text to change that pattern before the end of the decade.

Some, such as the Brussels think tank Bruegel, have **proposed** rebuilding the abandoned liability part of the EU AI framework. But there are no signs at present of this or any similar proposal gaining momentum.

Meanwhile, European member states are writing their own rules. Italy's **Law No. 132/2025**, approved in September 2025, is the first comprehensive national AI law in the European Union. It layers domestic criminal provisions and sectoral rules on top of the EU AI Act. This is exactly the outcome the European Union's harmonization machinery exists to prevent. The same cross-border AI deployment now faces different liability rules in different member states.

The United Kingdom tells the same story from outside the bloc. The Ada Lovelace Institute **studied UK AI liability** in December 2025. It found that UK law mostly does not treat AI as a "product" under the **Consumer Protection Act 1987**. AI systems are typically classified as services or "digital content" instead. A UK claimant injured by an AI system must therefore prove negligence—against a standard of care that has not yet crystallized in the AI industry. The **Automated Vehicles Act 2024** imposes strict manufacturer liability for cars in self-driving mode. Almost nothing else gets that treatment. In practice, UK liability allocation is governed by vendor contracts. The canonical example, per the Ada Lovelace Institute, is **OpenAI's standard business terms**, which cap the company's liability at 12 months of customer fees.

> *The underlying dynamic is identical. The public institutions are not setting the terms, so the private ones are.*

The point of this detour is not to argue that European AI policy is failing relative to American AI policy. It is to point out that the two systems are arriving at the same destination by different roads. In both jurisdictions, legislative liability allocation has stalled or been withdrawn. In both, market mechanisms are setting the de facto rules. In the United States, that means state-approved insurance exclusions and ISO endorsement language. In the European Union, it means national tort fragmentation and vendor-contract liability caps. The American story is faster and more visible because state insurance commissioners move faster than European Commission directives. The underlying dynamic is identical. The public institutions are not setting the terms, so the private ones are.

Two implications follow. The analytic implication is that the U.S. insurance retreat is not a product of distinctively American deregulatory politics. Europe wrote the world's most ambitious AI statute and still ended up with AI liability allocated by vendor contract. The policy implication is that the recommendations in Section 11 cannot be borrowed from a European model because no functioning model exists from which to borrow. What does exist—on a different continent, and on the opposite side of the policy ledger from both Brussels and Washington—is the subject of Section 10.

Section 10: The China Contrast: Insurance as Accelerator

On March 2, 2026, China's Ministry of Science and Technology, the National Financial Regulatory Administration, the Ministry of Industry and Information Technology, and the China National Intellectual Property Administration **jointly issued** 20 guidelines on what Beijing has begun calling

"sci-tech insurance." The guidelines name AI as a priority sector for specialized insurance products and dedicated risk-reserve systems, alongside integrated circuits, quantum technology, biomanufacturing, hydrogen energy and nuclear fusion, brain-computer interfaces, and embodied intelligence. The instrument was published in advance of the National People's Congress, against the backdrop of the broader sci-tech self-reliance push that Xi Jinping has named the central organizing concept of Chinese industrial policy.

The framing in the joint document is noteworthy: The guidelines call on the Chinese insurance industry to "fully leverage" its role as "an economic shock absorber and social stabilizer" in service of accelerating tech-sector adoption. These verbs are deliberately accelerationist. Insurance is being framed as an instrument that catalyzes the diffusion of high-priority technologies, not as a brake that adds costs and forces restraint. The American conversation, as Section 4 documented, is moving in the opposite direction: State insurance commissioners are approving more than 80 percent of carrier requests to strip AI from corporate policies, and the practical effect is a slow contraction of what enterprises can do with AI without bearing the residual liability themselves. Beijing is using the same instrument to produce the opposite outcome.

The concrete mechanisms in the Chinese guidelines suggest the policy direction is more than rhetorical. The framework calls for a coordinated national risk-reserve system supporting specialized insurance products in priority sectors, with three clusters, Beijing-Tianjin-Hebei, the Yangtze River Delta, and the Greater Bay Area designated as initial pilot zones. The People's Insurance Company of China (PICC), the country's largest property and casualty insurer and a state-controlled entity, has already begun underwriting standalone generative AI insurance covering IP, portrait, and reputational infringement claims arising from AI-generated content. Several Chinese carriers have piloted parallel products in the past two years, but the March 2026 guidelines move the activity from carrier discretion to coordinated state policy.

The growth ratios are striking even with the caveats appropriate to Chinese insurance statistics. Tech-insurance premium volume is **reported** to have grown roughly 44 percent year-on-year in 2025, against an industry-wide growth rate of approximately 7.4 percent—a six-fold differential between the priority sector and general lines.

> *Washington has not articulated a federal position on whether AI insurance capacity should expand or contract; Beijing has—and has organized the four ministries with the relevant authority around expanding it.*

The contrast with the United States is instructive precisely because it does not run on a simple ideological axis. Washington has not articulated a federal position on whether AI insurance capacity should expand or contract; Beijing has—and has organized the four ministries with the relevant authority around expanding it.

That difference matters for the AI race. U.S.-China competition in AI is usually described as a contest of models, chips, and deployment ecosystems. It is also a contest of who can build the most effective network of incentives to accelerate AI adoption and improve AI governance. That demands a focus on liability and on insurance. Which country's enterprises can deploy AI at scale without carrying the residual risk themselves? Which country's smaller firms can afford to participate at all? In the United States, the answer increasingly depends on a young specialty insurance market. Its limits top out at $50 million per insured. Most of it still does not cover users of third-party foundation models, including most small businesses in America. If the broader carrier retreat proceeds at the next renewal cycle, that thin market is all that stands between small American enterprises and uninsured AI adoption. Chinese firms face no such squeeze. Beijing is building a national risk-reserve system, backed by state-controlled carriers and provincial pilots, designed to prevent one. The constraint is asymmetric. The asymmetry runs in the direction Beijing intended.

None of this implies that the Chinese system is one Americans should want to import. The Chinese insurance industry operates under conditions—state-controlled major carriers, capital controls, limited claims litigation exposure—that the United States does not share and would not want. The point is narrower. Beijing has decided that insurance is a lever of AI industrial policy and has organized state action around expanding capacity in priority sectors. Washington has not decided. State insurance commissioners and ISO endorsement drafters have decided in Washington's place, and they have decided to contract capacity. That is the choice this paper's analytic sections are designed to surface and the recommendations in Section 11 are designed to make explicit.

Section 11: What Governments Should Do

Sections 3 through 10 describe a structural market failure with an unintended regulatory outcome. Carriers are pulling back from AI coverage on defensible actuarial logic. State insurance commissioners are approving the pullback within their existing authorities. The cumulative result is an AI deployment regime that no elected policymaker chose. This section proposes four recommendations to convert that drift into a deliberate policy. Each recommendation is tied to an existing institutional template. Each can be adopted on its own. None requires importing the European compliance model or the Chinese state-carrier model.

The four recommendations also map directly onto the insurability failures diagnosed in Section 3, and it is worth being explicit about which repairs which. The incident database in Recommendation 1 addresses the actuarial criteria—loss frequency and average loss amount—by producing the base-rate data that no single carrier can generate alone. The federal backstop in Recommendation 3 addresses maximum possible loss and, by extension, the acceptable-coverage-limits problem that keeps specialty products capped near $50 million. The verification and standards architecture in Recommendation 4 attacks information asymmetry, the failure this paper regards as the most fundamental of the three red verdicts. Adequate premiums are not addressed directly because they do not need to be: Premium adequacy follows arithmetically once severity, frequency, and verification improve. The two societal criteria—consistency with public policy and legal permissibility—belong to courts and legislatures rather than to any insurance instrument, and no recommendation here pretends otherwise. In short, the package is not a list of good ideas that happens to involve insurance. It is a targeted repair kit for the specific failures that make AI hard to insure.

One objection should be addressed first. Kenneth Abraham and Catherine Sharkey, two of the country's leading insurance law scholars, **argued in May 2026** that concerns about AI insurability "will prove to be either exaggerated or unwarranted." They point out that existing policies are already paying some AI claims and argue that a dedicated market will develop just as cyber insurance did. Recent market behavior gives these arguments some support. CFC, Coalition, and AIUC began writing affirmative AI coverage in 2025 and 2026 without any help from Washington (though the maximum coverage amounts were relatively small). The recommendations below, however, remain appealing even if one agrees with the objection. Three of the four recommendations are about strengthening information and verification infrastructure, not mandates. Their purpose is to make the market that Abraham and Sharkey are expecting arrive years, perhaps decades, sooner than would otherwise be the case. The fourth recommendation serves as a contingent backstop. If the optimists are right, it would never trigger and thus would cost nothing.

- **Recommendation 1: Build the AI incident database nearly all stakeholders agree is needed.**

  One of the largest barriers to AI insurability is an absence of loss data. The Lloyd's Market Association survey **put it plainly**: Insurers "are remote to insureds' use of AI," and little usage data is collected during underwriting. No single carrier can fix that. However, a shared, industry-wide incident database could.

  A **July 2026 industry report** (hereafter referred to as "the AIUC coalition report") written by 37 co-authors from AIUC, Aon, QBE, and Generali, along with researchers from Anthropic, OpenAI, and the RAND Corporation, recommends a voluntary anonymized AI incident database hosted at NIST, modeled on the Federal Aviation Administration's **Aviation Safety Reporting System** (ASRS). Even Daniel Schwarcz and Josephine Wolff, two of the sharpest academic skeptics of insurance as an AI regulator, reach the same conclusion from the opposite direction. Their **study** of the cyber insurance record concludes that regulators should mandate standardized AI incident data collection rather than wait for insurers to generate it. In Brussels, the think tank Bruegel has proposed a **public EU incident registry** and framed it explicitly as insurance market infrastructure. Illinois now requires frontier AI developers to report critical safety incidents within 72 hours under **a law** signed in July 2026. The industry coalition, its critics, Brussels, and a state legislature have converged on the same recommendation. That is a remarkable consensus and one where federal government intervention could improve the market's ability to effectively function and do so for minimal cost.

  The mechanism should run on two tracks because the underlying data problem has two halves: routine incidents that firms will share voluntarily if sharing is safe, and severe incidents that at least some firms will never disclose unless required to do so.

  The voluntary track requires no legislation. NIST would establish an anonymized AI incident database modeled on the **Aviation Safety Reporting System** (ASRS), which NASA has operated on the Federal Aviation Administration's behalf since 1976. The ASRS model works because of a specific design choice: Reports are stripped of identifying information, and filing a report confers some **limited protection** from enforcement action, which is why pilots and air traffic controllers voluntarily submit **tens of thousands** of near-miss reports every year that no regulator would otherwise see. The same bargain would work for AI. Companies that report incidents would

receive confidentiality and a safe harbor, and the market would receive the loss-frequency data whose absence is, as Section 3 documented, one of the **largest barriers** to AI insurability. Once incidents accumulate by sector and use case, actuaries can begin estimating base rates and (as appropriate) offering a priced premium for the coverage.

The mandatory track addresses what voluntary reporting will struggle to surface: severe incidents in healthcare and critical infrastructure, where the legal stakes of disclosure are highest and the public interest in disclosure is greatest. This track likely requires a legal statute. Legislation can also provide what interagency improvisation cannot: protection of submitted reports from discovery and Freedom of Information Act (FOIA) requests, a defined incident threshold, and a defined reporting timeline. Congress would not be designing from scratch. Illinois' new **frontier AI law** already requires critical safety incidents to be reported within **72 hours**, and **the Obernolte-Trahan** discussion draft includes federal incident reporting provisions.

Two further points argue for building the incident reporting infrastructure first. The first is that the security concerns that complicate cyber incident reporting mostly do not apply to AI. Disclosing an unpatched software vulnerability gives attackers a roadmap, which is why cyber information-sharing regimes are wrapped in protections and why many companies hesitate to use them anyway. Disclosing that an AI system hallucinated a medical code or approved a fraudulent invoice gives attackers nothing comparable. In this respect, AI incident data has more in common with aviation near-miss data or medical malpractice data than with cybersecurity data, and it can be pooled with fewer safeguards and less industry resistance. The second point is that the other three recommendations all consume what this one produces. The working group in Recommendation 2 cannot judge whether the exclusion templates are too broad without knowing which AI perils actually generate losses. The federal backstop in Recommendation 3 cannot be responsibly calibrated without loss data; as noted above, a backstop authorized before that data exists is a backstop authorized in the dark. And the verification reports that independent verification organizations generate under Recommendation 4 need somewhere to aggregate. Incident reporting is the cheapest of the four recommendations, the most aligned with existing authorities, and the one on which the other three depend. That is why it should come first.

- **Recommendation 2: Convene a federal-state working group before the exclusion templates harden.**

The second recommendation requires no new statute. **More than 60** property and casualty insurance groups have filed to adopt AI exclusions. Most filings use the **three ISO endorsement templates** that became available in January 2026. State approval patterns **vary widely**, and a second wave is coming. Verisk confirmed in July that it is weighing **new exclusions** for agentic AI. The National Association of Insurance Commissioners has standing authority to convene state commissioners on questions that cross state lines. It should convene them now, with input from the Treasury Department's Federal Insurance Office, the Commerce Department, and the Center for AI Standards and Innovation at NIST.

The working group should produce three outputs.

The first is a recommendation on how broad the exclusion templates should be. This matters because the broadest filed wording makes coverage depend on accusations rather than facts. W.R. Berkley's endorsement **bars coverage** for any claim involving "any actual or alleged use" of AI. Under that wording, a plaintiff's lawyer who mentions AI in a complaint can strip the defendant's coverage, whether or not AI had anything to do with the harm. That is excessive. A model recommendation would give commissioners a common yardstick for the filings already on their desks, in place of 50 separate judgment calls. And timing is the point: If carve-outs are warranted for perils such as bodily injury, privacy breaches, and employment discrimination, the group should say so before the agentic AI templates arrive and the same overbroad language hardens into a second generation of forms.

The second output of the working group should be a documented national map of the coverage gap the templates leave behind. No institution currently has one, not even the Treasury Department's Federal Insurance Office. Each commissioner sees only the filings in their own state, so the aggregate picture of what American businesses can and cannot insure exists nowhere. That is a strange condition for a question this paper argues is industrial policy. The map is cheap to build, since it compiles approved filings the states already possess. It is also the necessary predicate for Recommendation 3. Congress cannot scope a backstop for uninsurable losses, or decide whether one is needed at all, without knowing which losses are actually uninsurable.

The third output is a deadline for carriers to declare, line by line, whether AI risk is affirmatively covered, silently covered, or excluded. The goal of this is to eliminate illusory coverage. A business with silent AI cover believes it is protected until the claim arrives and the answer is decided in court. In February 2026, a Delaware judge ruled that more than 20 of Meta's insurers **owed no duty** to defend thousands of social media addiction claims pleaded as negligent design. That is what silence looks like when it is tested. Silence is also unpriceable because carriers cannot reserve against exposure they have not acknowledged. There is a proven approach to fixing this: Beginning in July 2019, Lloyd's **directed every syndicate** to state whether each policy affirms or excludes cyber coverage, phased in across all lines from January 2020. A declaration mandate will produce some new exclusions, as it did for cyber. But it converts hidden gaps into visible ones, and visible gaps generally get priced and filled. Buyers who know they are unacceptably uncovered demand affirmative products. Carriers that have stated their positions can price them. That sequence, clarity first and a standalone market second, is how cyber insurance became a real market. The **AIUC coalition** report wisely recommended the same treatment for AI.

- **Recommendation 3: Prepare a layered federal backstop for catastrophic AI losses, and stage it.**

  A single failure in a widely deployed foundation model could produce correlated losses across thousands of insureds simultaneously, a risk structure that (structurally) resembles terrorism, pandemic, and nuclear risk more than conventional casualty exposure. For each of those categories, the private insurance market proved willing to write coverage only after a government agreed to absorb the catastrophic tail. The **Price-Anderson Act of 1957** established a federal

backstop for nuclear operator liability after the private market declined to insure commercial nuclear power on its own. **Pool Re** was created in 1993 to keep London insurable against terrorism after the Provisional Irish Republican Army bombing campaign drove commercial reinsurers from the market. The **Terrorism Risk Insurance Act of 2002** did the same for the United States after the September 11 attacks. Pool Re was **extended in 2018** to cover physical damage from cyber terrorism, and TRIA today applies to **cyberattacks** that are certified as acts of terrorism. Thus, the precedent for adapting a catastrophic backstop to a digital risk already exists.

The **AIUC coalition report**, whose authors broadly favor government involvement, warns that TRIA's structure operates in practice as a subsidy and cites evidence that it has generated moral hazard. The better structural template is **Price-Anderson**, which layers the risk: Each nuclear operator carries a private retention, the industry as a whole funds a second layer through retrospective assessments, and the federal government stands behind both. The AIUC report sketches what an AI version **might look like**, with roughly $500 million in private retention and an industry-funded pool in the low tens of billions. The precise figures matter less than the sequencing principle. Federal capacity should sit above industry capital rather than in place of it, permitting carriers and developers to retain a financial stake in preventing the losses the backstop exists to absorb.

The statutory drafting matters. The cyber-line litigation around the "act of war" exclusion in Merck & Co. v. ACE American Insurance Co. is the cautionary tale: a roughly $1.4 billion coverage dispute arising from the 2017 NotPetya attack that was **affirmed on appeal in 2023** and **settled in January 2024**, only the day before the New Jersey Supreme Court was to hear argument. Statutory clarity on what constitutes a certifiable AI catastrophic event must be specified ex ante: a defined category of foundation model failure, a defined loss-aggregation threshold, and a defined certification authority, which in **TRIA's case** is the secretary of the treasury. Without that clarity, litigation over the trigger will consume the program's first decade and the mechanism would likely fail when it is needed.

- **Recommendation 4: License independent verification organizations—and accelerate the standards they verify against.**

Information asymmetry is the insurability failure that policy can repair most directly. Insurers cannot see how a customer actually uses AI. They cannot price moral hazard and adverse selection without verification. A relevant federal precedent here is the **SAFETY Act of 2002**, which grants liability protections to antiterrorism technologies that pass government vetting. The AI version is the independent verification organization (IVO), an idea **developed** by the nonprofit Fathom (which sponsored this research project). As designed, an IVO is a licensed, expert-led body that verifies AI systems against measurable risk standards. It is independent of the companies it audits. Its verifications carry defined legal consequences. In the past year, the model has moved from **white paper to statute** on three tracks at once: the private market, the states, and now Congress.

The private track is already operating. AIUC conditions its insurance coverage on certification against **AIUC-1**, an auditable standard for AI agent safety and reliability, checked by quarterly third-party evaluations. ElevenLabs became a **named policyholder** in February 2026. This is verification purchased voluntarily because it unlocks insurance.

Verification, however, is only half of the standards layer. An IVO needs something auditable to verify against, and the diagnosis in Section 3 is at bottom a standards gap: There is not yet a widely accepted methodology for assessing the risk of a deployed AI system in terms an underwriter can price. AIUC-1 is the first insurance-linked attempt to fill that gap, and it will not be the last—several industry consortia and trade associations are now developing system-level AI risk assessment standards aimed at the same target. Federal policy should treat these emerging standards the way it has treated the National Fire Protection Association's model codes since the nineteenth century: as privately drafted raw material that public institutions harmonize, reference, and give legal effect. CAISI and NIST should convene the organizations developing AI risk assessment standards, drive convergence on shared definitions and evaluation protocols, and publish the result as a reference framework that IVOs, underwriters, and state regulators can adopt wholesale. Standards development is the cheapest intervention proposed in this paper, and, as Section 2 documented, it is the mechanism by which insurance markets have disciplined every prior technology: The standard comes first, and the market forms around it.

The state track advanced in the spring. Virginia moved first, directing a **formal study** of the IVO model in April. Connecticut enacted it. Governor Ned Lamont signed **House Bill 5222** on June 2, now Public Act 26-100, after both chambers **passed it** with bipartisan supermajorities. Section 47 authorizes the state Department of Consumer Protection to license up to five IVOs in a multiyear pilot beginning July 1, 2027. The division of labor is the important design feature. The state sets the risk mitigation and harm prevention outcomes. The licensed IVOs develop the technical criteria and verify whether specific AI deployments meet them. Participation is voluntary. The incentive is legal: Verified deployments gain defined evidentiary weight in civil litigation. The pilot includes a built-in evaluation by the University of Connecticut, with a statutory mandate to recommend whether the program becomes permanent and whether Connecticut should recognize other states' IVO regimes. State Senator James Maroney, the bill's champion, built it directly on **Fathom's framework**.

California, the largest state insurance market in the United States, is moving on a parallel track. **Senate Bill 813**, introduced by State Senator Jerry McNerney, would establish licensed private certifying bodies—structurally similar to IVOs—whose certification of an AI system would confer a liability safe harbor, while Assemblymember Rebecca Bauer-Kahan's companion **AB 1405** would create an enrollment and oversight regime for third-party AI auditors; the two lawmakers have **presented the bills** as a package. Whatever the bills' fate in the current session, California adds an ingredient no other state can: an elected insurance commissioner regulating the largest insurance market in the country. A commissioner who encourages carriers to attach underwriting credit to certified deployments—lower premiums, or restored coverage, for systems verified against a recognized standard—would do more to create market demand for verification than any audit mandate now pending in any legislature. That is the standards-to-insurance flywheel this paper has described, and it could be started by a single state officeholder using existing authorities.

Ohio's **House Bill 628**, introduced in December 2025, spells out the fullest version of the legal architecture. The state attorney general licenses the IVOs. Verification produces a rebuttable presumption against tort liability for harms within the verified risk category. The presumption can be defeated only by clear and convincing evidence of intentional misconduct, material

misrepresentation to the IVO, or concealment of material risks discovered after verification. IVO personnel may not hold equity in, or employment with, the companies they verify. Records must be kept for 10 years. As of early August 2026, the Ohio bill remains in committee. Even so, it is the reference text for how a mature IVO statute should look.

The contrast with the other 2026 state AI laws is instructive. **Illinois** and **New York** now require frontier AI developers to obtain third-party audits, but neither state licenses the auditors, sets standards for their independence, or attaches any legal consequence to the results. Without those features, an audit requirement risks becoming a compliance exercise in which developers select their own reviewers and file the resulting paperwork. What distinguishes the Connecticut and Ohio approach is that it builds the auditing institution itself: The state licenses the verifiers, polices their independence, and gives their verifications defined weight in civil litigation.

Beyond the state-level movement on IVOs, there is now draft bipartisan congressional legislation. On June 4, 2026, Representatives Jay Obernolte and Lori Trahan released the **Great American AI Act**, a **269-page bipartisan discussion draft** joined by four House colleagues from both parties. It is the first comprehensive federal AI governance framework proposed in Congress, and IVOs are its enforcement backbone. Under the draft, the Center for AI Standards and Innovation **licenses** the IVOs and is funded at **$100 million per year**. Large frontier developers, defined as those with more than $500 million in annual revenue building cutting-edge models, must retain a licensed IVO within one year of enactment and every six months after that. The IVO verifies compliance, assesses the adequacy of the developer's safety framework and risk monitoring, and reports any failure, deficiency, or material weakness in controls. Developers must give their IVO timely access to unredacted materials, subject to trade-secret protections. CAISI's director can order additional audits after a critical safety incident or a major model change. Violations carry fines of up to **$1 million** per violation per day.

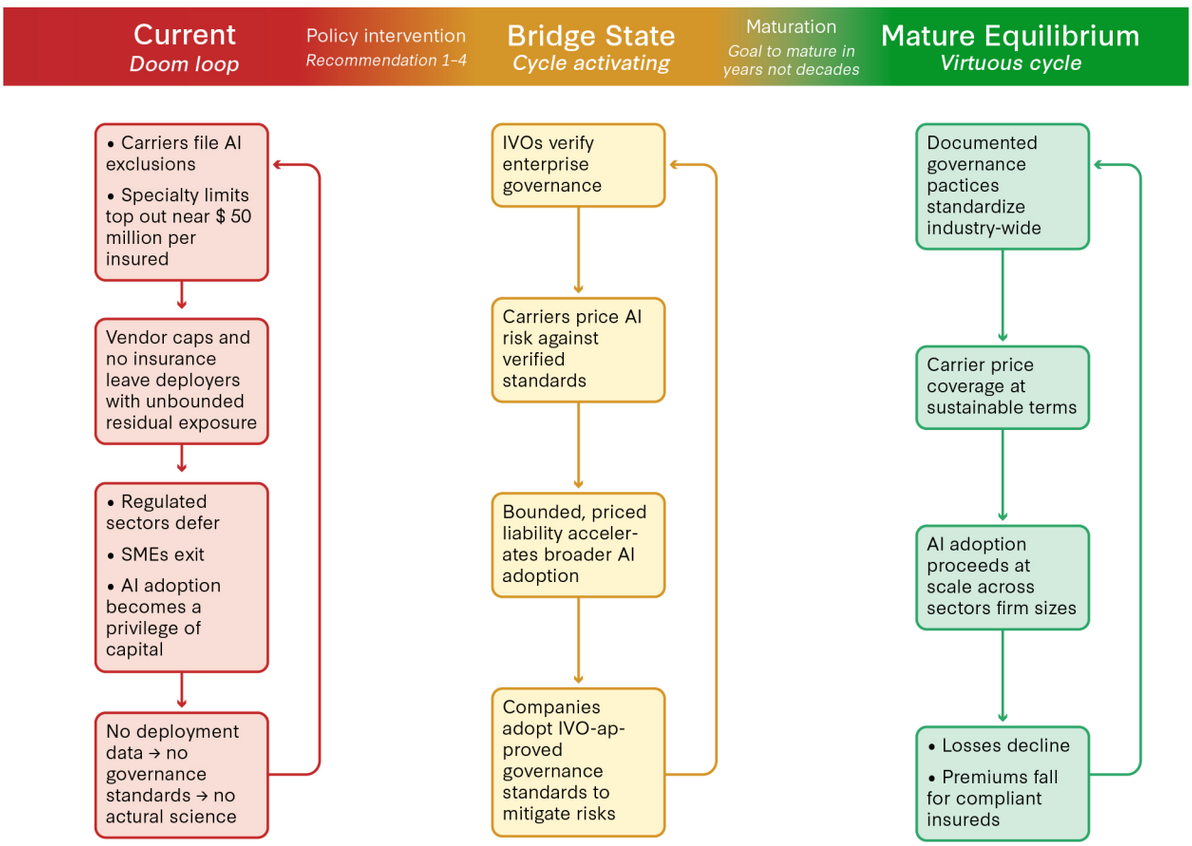
The federal draft does contain one provision that should give IVO supporters pause. For three years, it would preempt state laws that specifically regulate the development of AI models while preserving state authority over how AI systems are used and deployed. Where the line between development and use will ultimately fall is not yet settled, and no published analysis has yet mapped the existing state laws against it. The developer obligations that Illinois and New York enacted this year would appear to sit close to the line. Connecticut's pilot seems a harder target under any interpretation, since it imposes no requirements on developers and instead operates as a voluntary verification program that confers legal benefits on participants.

The IVO model does face one serious published objection. Gabriel Weil has argued that allowing AI developers to select and pay their own verifiers will recreate the dynamic that undermined the credit rating agencies before the 2007–08 financial crisis, in which issuers shopped among raters until they found a cooperative one. The failure mode Weil describes is real, but the IVO statutes were drafted with it in mind, and their design differs from the rating agency precedent in the ways that matter. The rating agencies were paid by the issuers they rated and supervised by no one. IVOs, by contrast, operate under a public license, whether from Connecticut's Department of Consumer Protection, Ohio's attorney general, or CAISI under the federal draft, and both state bills prohibit IVO personnel from holding equity in or employment with the companies they verify. Ohio's bill goes further, stripping the liability presumption from any verification obtained through

misrepresentation. Nor would IVOs displace the capital-backed scrutiny that Weil prefers. An insurer that relies on an IVO verification still has its own money behind the policy, which means verification would supplement underwriting discipline rather than substitute for it.

A final design question deserves an explicit answer: whether IVOs are transitional scaffolding for an immature market or a permanent feature of a mature one. The information asymmetry diagnosis in Section 3 implies the latter. An insured will always know more about its AI deployment than its carrier, just as a building owner will always know more about its wiring than its fire insurer—which is why the institutions that verify against fire codes and crash-test the vehicle fleet did not dissolve once those insurance markets matured. Underwriters Laboratories is more than 130 years old; the Insurance Institute for Highway Safety is more than 65. If the AI insurance market matures along the path this paper recommends, IVOs will not fade away. They will become its permanent verification infrastructure—as unremarkable, and as load-bearing, as the fire inspector.

Figure 1: A Potential Pathway for AI Insurance



Source: CSIS.

Section 12: Conclusion: Insurance Policy Is AI Industrial Policy

This paper opened with oil tankers because the 2012 episode illustrates a form of power that is easy to overlook until the moment it is used. The United States did not need to board ships or blockade ports to slash Iranian oil exports because cutting off access to insurance accomplished the same thing: An oil

tanker that cannot be insured is an oil tanker that cannot dock. The same power—the ability to grant or withhold the coverage that permits commercial activity—is now being exercised over American AI deployment, and it is being exercised without anyone in particular having decided to exercise it.

None of the actors in this story is behaving unreasonably. Carriers are responding to a litigation curve that their actuaries cannot yet price, state insurance commissioners are approving exclusion filings that fall squarely within their existing authorities, and frontier labs are capping their contractual liability because no insurance market will absorb the exposure they would otherwise be retaining. The trouble lies in what all of these individually reasonable decisions add up to: a de facto regulatory regime for American AI that is harsher in its practical effect than anything Congress has seriously contemplated, and that is being written into endorsement language and reinsurance treaty wordings that in all likelihood no elected official will ever read.

The four recommendations offered in this paper are deliberately modest relative to the scale of that problem, and they are modest in a particular way. Rather than proposing that the federal government regulate AI at large, each recommendation borrows an institutional template that already exists—such as the NAIC's convening role, the incident-reporting bargain of the Aviation Safety Reporting System, the layered retention structure of Price-Anderson, and the certification logic of the SAFETY Act—and applies it to a specific, named failure of insurability. The sequencing matters as much as the substance. The incident database should come first because every other recommendation consumes the data it produces, and the federal backstop can come last because, if the optimists about AI insurability turn out to be right, it will never trigger and will never cost the taxpayer anything.

What separates the United States from China on this question is not capability but decisiveness. Beijing has concluded that insurance is an instrument for accelerating AI adoption and has organized four ministries around that conclusion, while Washington has yet to conclude anything at all, which means that the terms of the American AI revolution are currently being set in 50 state insurance filings that almost no one in Washington is reading. There is still time to choose deliberately what is now happening by default, and the stakes of that choice are the ones this paper began with: For most American businesses, an activity that cannot be insured is an activity that is effectively prohibited. That rule determined which tankers could sail in 2012, and it will determine what American businesses can do with AI in 2027. ■

Gregory C. Allen *is a former senior adviser with the Wadhwani AI Center at the Center for Strategic and International Studies in Washington, D.C.*

This report is made possible by generous support from Fathom.