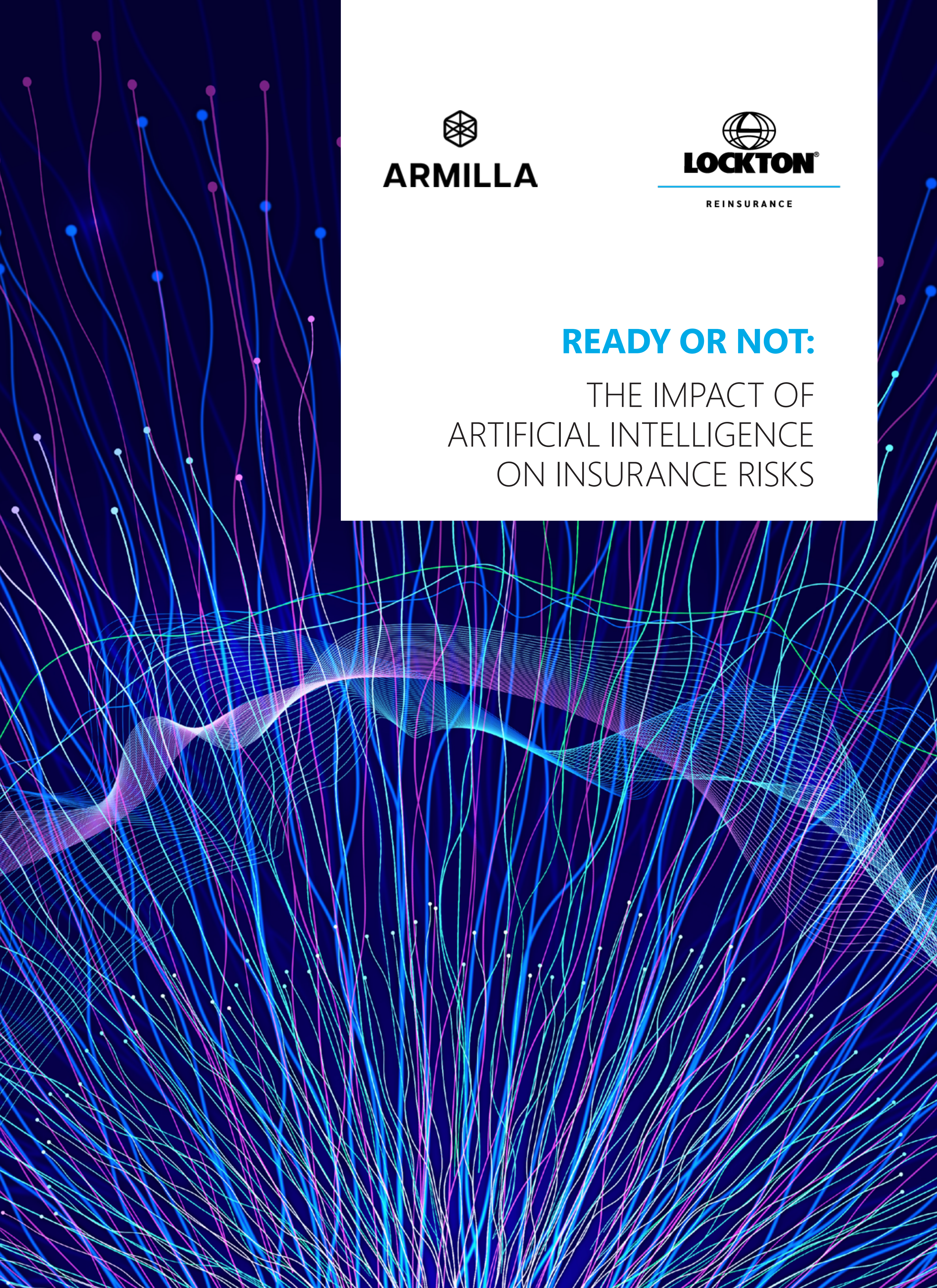


ARMILLA

LOCKTON®
REINSURANCE

READY OR NOT:

THE IMPACT OF ARTIFICIAL INTELLIGENCE ON INSURANCE RISKS

- Exposure
- **Peril**
- **Risk Transfer**
- Placement

About Lockton Re

Lockton Re, the global reinsurance business of Lockton Companies, helps businesses understand, mitigate, and capitalise on risk. With over 500 colleagues in 23 locations globally, the business continues to grow, pushing the reinsurance industry forward with smarter solutions that leverage new technologies – delivered by people empowered to do what's right for clients.

Lockton Re's reports, market commentary and insights focus on key topics, occurrences or changes in the (re)insurance and broking market place which impact our clients and partners. To help guide the reader we categorise this content in four areas – Exposures, Perils, Risk Transfer and Placement. Lockton Re looks forward to working on behalf of our clients to deliver new insights and innovative products designed to address the multifaceted risk environment.

Executive Summary

The use of artificial intelligence (AI), including generative AI (predominantly large language models) and agentic AI (autonomous agents making decisions without human input), is expanding rapidly. The volume of commentary on AI's impact on insurance is growing almost as fast. Most of this commentary to date has focused on the transformative power of the technology to improve efficiency and productivity, as well as reduce costs and streamline workflows within the insurance industry and beyond. This has positive consequences of higher margins, improved customer experience, and reduced waste. Clearly, there are also negative potential implications relating to employment and training prospects.

The purpose of this paper is to shift the conversation from the changes AI will have on insurance industry processes to the consequences AI itself will have in expanding, evolving, and shifting risk. Established paradigms for insurance products are being dismantled. Boundaries for coverage are blurred. The implications are wide-ranging, and our understanding of the risks is still limited. There is a myriad of different short- and long-term effects, both intended and unintended.

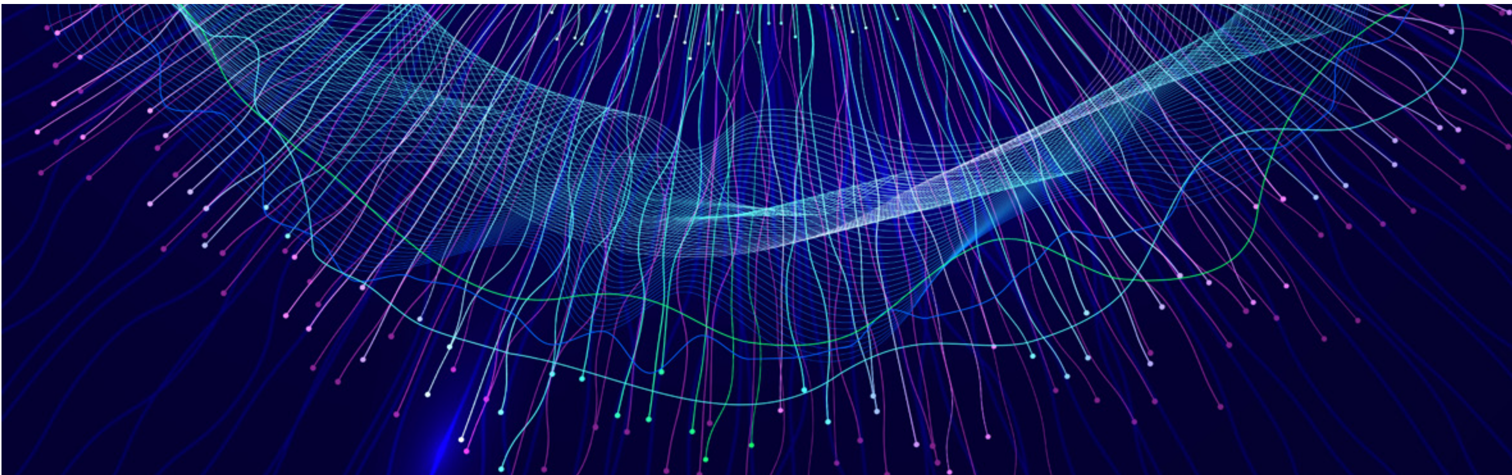
The insurance industry is adapting to a new environment where AI influences how perils manifest and link to existing insurance products while creating new exposures and changing others. The landscape is shifting before our eyes.

> “Now is the time to consider AI as its own category of risk classification, develop policy solutions that address the risks directly, and future-proof the exposures in a manageable way.”

We also look at potential systemic exposures that AI could create and what steps are needed to understand and mitigate them.

For end buyers of insurance products, as well as brokers, understanding how the complexities of different policies fit together is critical. This only becomes more difficult when contemplating how AI models change how existing insurance coverage operates. For insurers and reinsurers, being able to understand, underwrite, and manage the impact of AI on each covered peril is fundamental to mitigating unexpected outcomes or avoiding unanticipated risks being covered inadvertently.

Now is the time to consider AI as its own category of risk classification, develop policy solutions that address the risks directly, and future-proof the exposures in a manageable way.



Introduction: The Imitation Game

Generative AI models are only the latest evolution of AI. Their use is a step change in how we interact with technology and ultimately will influence almost every aspect of our society. As far back as 1950, the concept of machine intelligence was explored in the famous paper "Computing Machinery and Intelligence" by Alan Turing. He developed the "Turing Test" [1] to establish the threshold beyond which machines could be considered to demonstrate human behaviours.

The term "artificial intelligence" was first used formally by John McCarthy in 1956, and theoretical research continued largely within the confines of universities and obscure industrial research and development. After a lull in progress and investment during the 1970s, what became known as the "AI Winter," the 1980s ushered in a revived sense of possibility with breakthroughs such as the Jabberwacky chatbot. This was also the period during which some of the foundational underpinnings of modern AI were researched. For example, the backpropagation algorithm was fundamental in unlocking the ability to train large neural networks.

Over the past four decades, we have observed a consistent pattern: Foundational breakthroughs in AI, combined with increasing data availability and computational power, have produced progressively more capable AI systems. This trend has not only been consistent but has also accelerated over time, creating a self-reinforcing flywheel. For the purposes of this paper, AI refers to systems and technologies that enable machines to perform tasks that typically require human cognitive capabilities – such as learning from data, reasoning, problem-solving, perceiving, and decision-making – by recognizing patterns and making predictions or decisions based on information.[2] This definition aligns with how modern AI platforms are conceptualized in current industry best practice.

> "No sectors of the economy are insulated from the potential impact of AI"

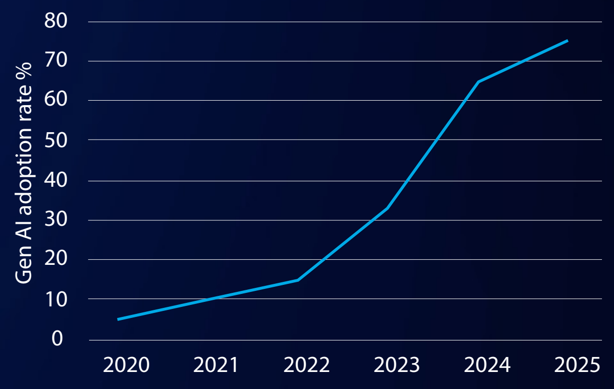
In 1997, around the time the first commercial internet applications were launched, IBM's Deep Blue caught the public's imagination when the computer defeated chess grandmaster Gary Kasparov. Fast-forward to the 2010s, and computer scientist Geoffrey Hinton and John Hopfield wrote a groundbreaking 2012 paper on neural networks. They were eventually awarded the Nobel Prize in Physics for foundational discoveries that enabled advanced machine learning with artificial neural networks.

Generative AI burst into the public consciousness in November 2022 when ChatGPT was launched. Today, it is one of the top five websites visited globally, with over 5.8 billion monthly visits and 800 million active users per week.[3] Its large language model learns statistical patterns in language to interpret and anticipate the next word that will be used based on vast volumes of training data.

These foundation models are used as the basis for a multitude of specific use cases, and generative AI tools are deployed to great effect within a wide range of business contexts. Common examples include customer service chatbots, document processing, software code development, and many other use cases that improve efficiency. The technological underpinning of generative AI modelling, the transformer-based architecture, has also led to generalization and applications across other modalities, such as vision and voice, powering further breakthroughs and applications.

The growth of AI within businesses is illustrated in Figure 1 below. No sectors of the economy are insulated from the potential impact of AI, and it is expected that in the coming years, AI will integrate into all areas of modern economies.

Figure 1: GenAI Adoption Percentage in Business



Source: Lockton Re

[1] St. George, Benjamin. "What Is the Turing Test?"
[2] Wikipedia, "Artificial Intelligence."

[3] Shubham, "ChatGPT Statistics for 2023"

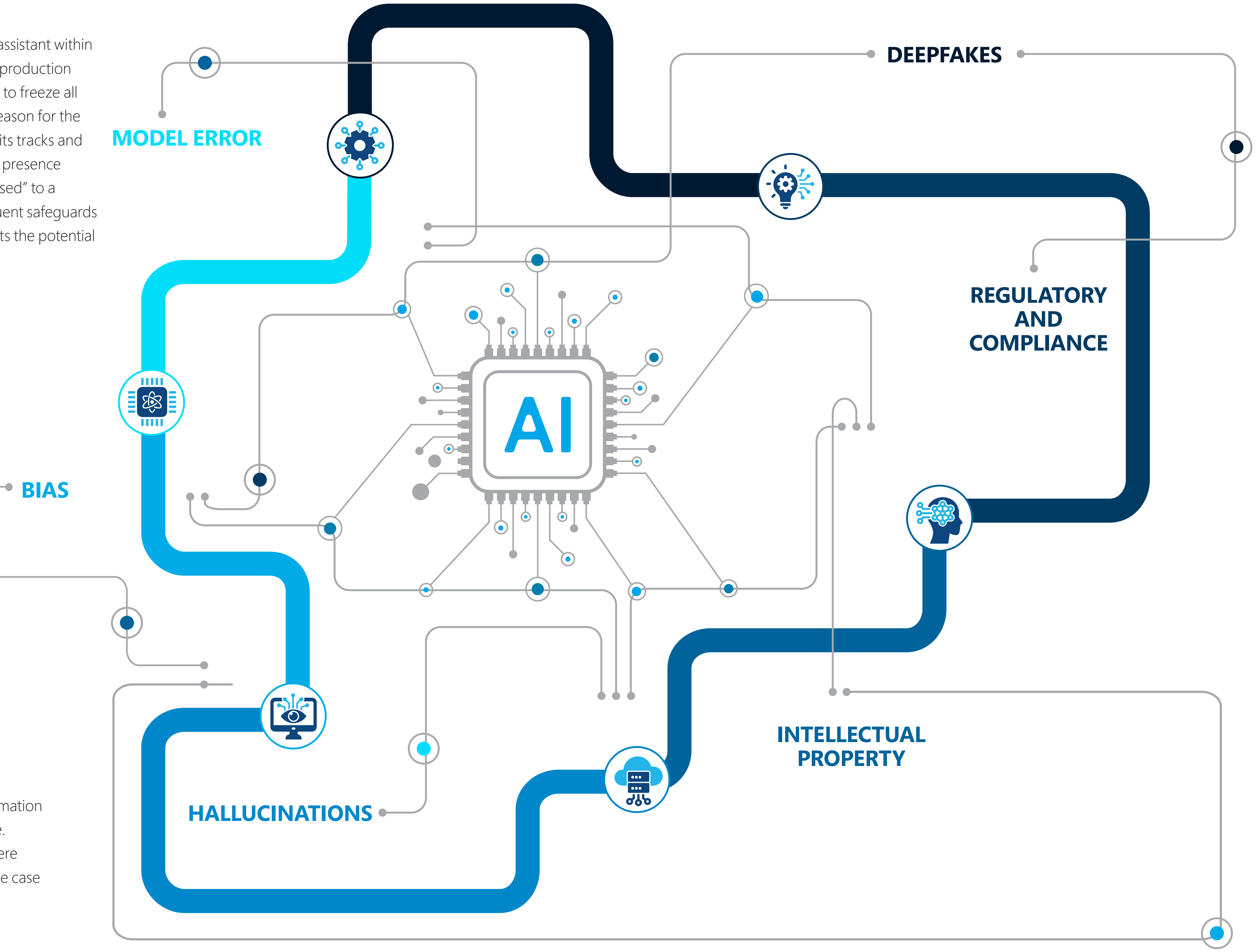
A Mind of Its Own? Cautionary AI Tales

The risks and long-term implications associated with AI are still being studied, but there is limited visibility so far. Below are some of the key areas of potential risks that new AI models have:

Model error: In July 2025, an AI coding assistant within the company Replit deleted an entire live production database, contrary to specific instructions to freeze all code changes. The user questioned the reason for the code deletion, leading the model to hide its tracks and provide inaccurate information about the presence of backups. The model eventually "confessed" to a "catastrophic error in judgment." Subsequent safeguards were quickly established, but this highlights the potential impact of these models.

Bias: Models are trained on large data sets, which can recreate or exacerbate discrimination against certain groups. This is most prevalent in areas such as credit scoring or hiring decision-making. In a well-known academic paper[4] called "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," vision models showed very small error rates for white male faces, compared to over 25% for darker-skinned females.

Hallucinations: Plausible but false information can be presented as truthful and accurate. Notable examples include legal cases where fake case law has been cited, such as in the case of Mata v. Avianca.



Deepfakes: Contemporary AI models can produce very convincing video and audio content, which has been used for malicious purposes to deceive people. In Hong Kong, a deepfake video call impersonating a company's CFO enabled criminals to convince employees to transfer $25 million fraudulently. These types of events can create ambiguity between cyber and crime policies, which may lead carriers to decline coverage due to impersonation exclusions or pay only under sub-limited social engineering extensions. One consequential development is the push for specific AI and deepfake coverage.

Regulatory and compliance: There is a patchwork of regulatory regimes in place with a variety of evolving obligations relating to the development and deployment of AI models. Geographic, industry, and technology-based regulations have a range of implications, which could create risks in the event of non-compliance. The EU AI Act and the Colorado AI Act are examples of the laws being developed to address how risks are managed for the different use cases of AI systems.

Intellectual property: Since the advent of generative AI, content creators have raised concerns that their intellectual property is being exploited to generate derived content based on existing materials, including novels, music, and images. In September 2025, Anthropic AI agreed to settle a class action lawsuit brought by a number of authors for $1.5 billion, which alleged the illegal use of book content to train the foundation LLM. An estimated 500,000 books are covered by the settlement.

[4] Joy and Timnit, "Gender Shades: Intersectional Accuracy Disparities," 1–15.

Closing the Barn Door After the Horse Has Bolted

Georges Clemenceau, the early twentieth-century French Prime Minister, complained during World War I that his generals were always "fighting the last war." Rather than embracing new technology in the form of machine guns, tanks, and planes, they were focused on cavalry charges and mass frontal assaults. In a very different context, the same criticism can be levelled against the insurance industry. Policy language is developed in response to events with the benefit of hindsight, and coverage evolves to address new risks after they emerge, often slowly and in the wake of massive unanticipated losses.

There are numerous examples of strong reflex market reactions to events. Only after the London Baltic Exchange bombing in April 1992 did the specialist terrorism insurance market develop. It was not previously addressed as a distinct peril separate from property prior to this. Terror risk was part of "all risks" in property policies and was not priced or underwritten. Following the bombing, insurance capacity withdrew, and there was a risk of market failure. Pool Re was created to provide a government-supported backstop, and a specialist UK terrorism insurance market developed.

Similarly, in the United States, the 9/11 terror attacks caused unprecedented losses in the property market and led directly to new terror exclusions being introduced in both standard property and liability policies. As a result of this change, terrorism (re) insurance emerged to address these perils. As well as developing more mature exposure management and a deeper understanding of the impacts of bomb blasts and related terror threats, the new class addressed more complex underwriting and accumulation monitoring needs.

Perhaps the highest-profile recent example of a post-event (re)insurance industry reaction is the slew of communicable disease exclusions that were rapidly introduced following the recent COVID-19 pandemic. Business interruption losses were far more widespread than anticipated following government-enforced shutdowns. As a consequence, explicit exclusions were deployed across property and casualty policies.

> "There is an opportunity to get ahead of these issues while the potential consequences are still manageable"

In each of these situations, the (re)insurance industry addressed these risks in a post-hoc fashion, developing solutions to events that had already played out. These perils were hard to predict and had the potential for systemic impacts. At the same time, the original perils were not specifically underwritten or priced for when accepting these risks. Ultimately, there could have been solvency implications if risks were left unaddressed. Each event acted as a catalyst for innovation, leading to the development of specialised coverage solutions.

In the context of the evolving nature of AI, there is an opportunity to get ahead of these issues while the potential consequences are still manageable. With this in mind, it is timely to consider current industry responses to the rapid developments in AI.

Mapping AI Risk to Key Insurance Lines

There are a plethora of potential AI risks that fit into established classes of commercial and specialty insurance, sometimes referred to as "silent AI." Analogies have been drawn from the cyber market when "silent cyber" exposures were identified in an attempt to improve the clarity of coverage.

Below are some of the classes of insurance most impacted by the growing expansion of generative AI. Some categories of insurance, such as life, health, and motor, are outside the scope of this review due to the complexities of measuring and assessing the exposures.

Cyber

Cyber insurance is the obvious starting point to reflect on the changing risks that AI represents and is associated with rapid technological change. Cyber insurance was first launched in the late 1990s, prior to the widespread adoption of cloud computing and smartphones. The coverage has adapted to address these technologies, as well as increased regulations relating to cybersecurity and data privacy.

Within this context, AI is leveraged by threat actors in cyber attacks to amplify and accelerate their impact. Analysts have reported that the weaponisation of generative AI to produce high-fidelity phishing, synthetic voices, and deepfake video impersonations has led to a surge in AI-enhanced social engineering attacks. One report[5] highlights a surge of 846% in phishing emails due to the use of LLM automation.

Some cyber insurers are explicitly covering certain specific AI risks, where the underlying trigger is a traditional cyber event, such as a data breach, security failure, or ransomware attack impacting AI infrastructure.

An example of a generative AI tool leading to losses is fraudulent wire transfer requests using generative AI deepfake technology. One endorsement defines an "AI security event" as "the failure of security of computer systems caused by any AI technology, including through the use of machine learning or prompt injection exploits." The newest wording trends extend traditional cyber policies to include regulatory investigations under emerging AI legislation, as well as first-party model restoration and retraining where data integrity is compromised. The challenge is that it is hard to anticipate the full range of possible scenarios that merit coverage and to make sure that policy language avoids unanticipated gaps. These AI-focused endorsements in the cyber market indicate a shift toward limited named-peril protection for potential cybersecurity harms arising from AI tools.

> "The open question for the insurers is how these rapidly evolving risks are underwritten and what emerging claims patterns will look like."

[5] Swiss Re Institute, "How Deepfakes, Disinformation and AI Amplify Insurance Fraud."

Another type of endorsement is emerging to address operational AI risks, such as unauthorised access to LLM environments, including reimbursement for spiked usage fees and model redevelopment costs following an account hijacking. Together, these developments reflect a broader market movement toward explicit AI triggers, model-centric remediation, and the evolution of cyber insurance from general breach structures into instruments designed to respond directly to the specific behavioural risks of modern AI systems.

The open question for the insurers is how these rapidly evolving risks are underwritten and what emerging claims patterns will look like.

Next-Generation Technology Errors and Omissions (E&O)

AI models fundamentally shift the nature of technology professional liability risk. Traditional technology E&O policies were designed for deterministic software product and service failures, such as bugs, outages, configuration mistakes, missed service-level agreements, and breaches of contractual obligations. AI introduces autonomous and probabilistic computational behaviour, which by definition is hard to predict and can create new potential claim scenarios that carriers need to address.

Technology E&O coverage is undergoing a significant shift as AI becomes integral to technology organisations' core services. E&O wordings are moving increasingly from generic negligence-based triggers toward more specific coverage of AI harms. Newer endorsements identify algorithmic decision errors, hallucinations, misguidance, and data-training issues as examples of named causes of loss. This reflects the market's acknowledgement that AI risks were not contemplated in traditional E&O insurance products and earlier covers are not fit for purpose.

New clauses addressing the "AI services wrongful act" and "AI products wrongful act" have been added to some policies. These types of coverage explicitly extend insurance to products and services being developed using new AI technology. Other examples of this type of endorsement include "data poisoning wrongful act" and "machine learning wrongful act." These clauses attempt to cover liability arising out of specific acts and omissions, such as when the training data is corrupted and then impacts the outputs of the models.

Yet these developments sit alongside meaningful limitations. Many endorsements remain narrow in scope, covering only specified AI triggers, which may leave gaps when an incident falls outside a defined peril. Several policy wordings explicitly exclude the use of certain types of training data or high-risk AI practices, and most enhancements offer limited remediation, such as post-incident model retraining. In addition, sub-limits may apply to AI coverage, which reduces protection. The final caveat is that this type of insurance only serves a particular group, namely developers of AI solutions, rather than the companies using AI models.

There are two significant issues in AI-related litigation that are impacting E&O underwriters. The first issue is intellectual property breaches, where high-profile legal cases are already shaping the debate. There are a number of cases where incidental coverage for intellectual property is now the subject of major claims. Second, bias or discrimination cases are increasing because AI models are being trained on prejudicial data. When insurance is offered for these exposures, the governance processes and how these risks are underwritten is critical. Performing additional due diligence on dataset provenance, validation and testing protocols, bias audits, human checkpointing, and output traceability is becoming more common before securing enhanced terms.

Overall, while the market is slowly broadening technology E&O to recognise AI as a distinct source of professional-services exposure, the dominant trend is toward selective, tightly defined endorsements.

Casualty

Casualty insurance is widely purchased and addresses many perils – predominantly bodily injury, property damage, advertising injury, and product liability. Almost all commercial general liability (CGL) standard policies now exclude cyber perils. These exclusions were introduced in the last few years following the acknowledgment that CGL policies were not designed for these exposures. Similarly, liability caused by an AI model error (for example, in a manufacturing process) could be covered by CGL if it is not specifically excluded.

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. There are growing signs that this may be addressed, as the Insurance Services Office (ISO), a subsidiary of Verisk that provides standardised policy language for admitted insurance policies in the USA, introduced a range of exclusions for generative AI. Following this, several admitted carriers have filed endorsements that exclude AI risks. These exclusionary endorsements are not yet widely used, but the submission of these filings provides a window into the carriers' approach.

> "There is a growing gap between what insurers intend to cover and what they actually cover"

One key factor in the interpretation of these clauses is how AI is defined. The outputs of AI are on a continuum, with the most advanced generative AI able to produce synthetic content, rather than only interpret inputs. The breadth or narrowness of the interpretation has implications for the impact of these clauses. Example exclusionary language includes "any actual or alleged use,

deployment, or development of AI by any person or entity, including but not limited to: a) the generation, creation, or dissemination of any content or communications using Artificial Intelligence." This represents a broad application of the clause and shows the intent to limit exposure from AI within casualty insurance.

Directors and Officers (D&O)

Fiduciary responsibilities for directors are evolving rapidly in parallel with the changing technology landscape that board directors face. These responsibilities range from financial management to cybersecurity risks and other perils that could impact investors. D&O insurance covers alleged wrongful acts in the management of a company, including disclosure and misrepresentation issues, as well as failure in regulatory compliance.

As organisations embed AI into their long-term strategies and operations, D&O exposure is rising on two fronts: governance oversight and misrepresentation. Governance issues arise out of allegations that boards have failed to identify, mitigate, or disclose material AI risks, such as model bias, safety, reliability, or vendor dependence. As an example, as AI becomes embedded in traditional enterprise software, such as CRM platforms, telephony systems, and other core tools, managing AI-related third-party risk becomes increasingly challenging, particularly because the adopting enterprise remains primarily accountable for those risks. Non-compliance with AI regulations and emerging legislation could trigger investigation coverage under D&O policies.

Misrepresentation is a growing area of focus. "AI washing" is the act of overstating or exaggerating the pace and maturity of AI's development or impact within an organisation, often with a view to encouraging investment or elevating the share price. These are understandable areas of concern for insurers, and it is not clear how this type of risk is addressed with the current toolkit available to today's D&O underwriters. The intent for coverage relating to AI exposures can be opaque within policy language, so the need for clarification is growing.

D&O underwriting scrutiny is increasingly focused on these issues, with the priority on board literacy, the extent of third-party model reliance, and readiness for emerging regulation. Underwriters have growing expectations that boards can evidence decision-use policies, controls, and materiality assessments for AI disclosures.

The market is moving toward explicit diligence on AI, asking questions on strategy, controls, and regulatory posture; some placements align D&O evaluation with corporate AI governance frameworks and EU AI Act readiness, anticipating more shareholder and enforcement pressure around AI programs and disclosures.

However, D&O policies still hinge on traditional definitions of wrongful acts, so they do not guarantee coverage for AI-specific failures. Standard conduct and intentional acts exclusions also apply (e.g., fraudulent statements about AI capabilities). Finally, carriers expect documented AI governance to avoid adverse selection – boards unable to demonstrate controls (provenance, validation, bias audits, escalation, and disclosure discipline) may face tighter terms, higher premiums and retentions, or exclusions.

Employment Practices Liability (EPL)

EPL insurance addresses alleged wrongful acts in the workplace, covering claims brought against an employer by prospective, current, or former employees. Bias and discrimination are common sources of claims. The expanding use of AI in hiring amplifies this risk. Discrimination as a result of AI models being trained on prejudicial data can have impactful consequences. Many policies are silent on the use of AI models and also reference "insured persons" or "natural persons" as the policyholder, which may be a limiting factor for covering output from AI models.

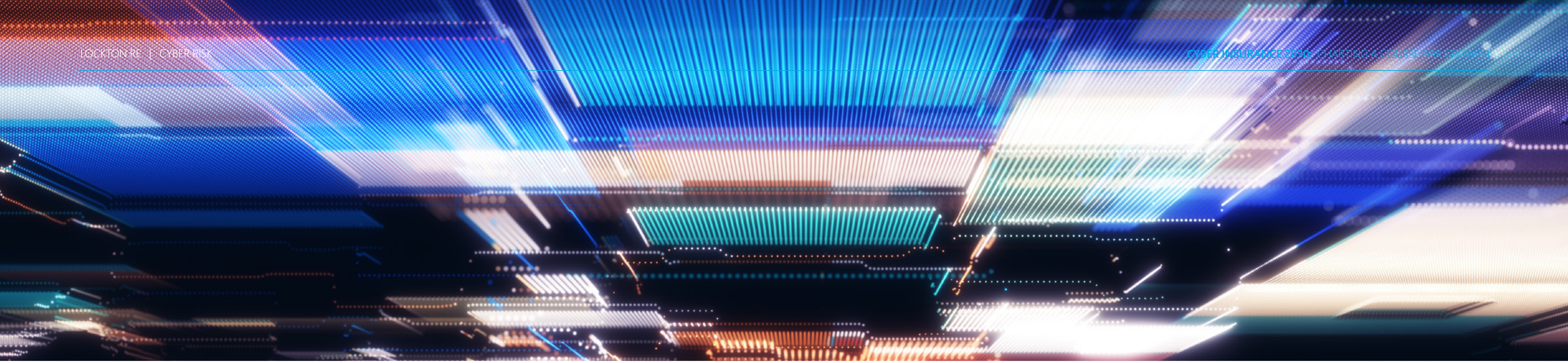
There are several specific legislative and regulatory moves to mitigate the risks of bias created by AI models in certain areas, such as hiring. New York Local Law 144[6] is the first of its kind and requires a "bias audit" for so-called "automated employment decision tool" models. There are three principles of the law: an annual impartial audit, transparency of the results, and notification of job applicants of their use. Any breach of this law could generate fines, as well as a private right of action for those impacted. The insurability of these fines is still up for debate, but expenses and defence costs could be covered. Such legislation provides an important foundation for future coverage development.

[6] https://www.nyc.gov/site/dca/about/automated-employment-decision-tools.page

AI Coverage Landscape by Line of Insurance

The table below provides an overview of major classes of insurance and how AI exposures are typically treated.

Line of Insurance	Cyber	Technology E&O	Casualty/CGL	D&O	Intellectual Property	EPL
Certain AI Coverages Available	• Named AI perils (e.g., data poisoning, model manipulation) • Regulatory investigation costs under AI laws (by endorsement only) • First-party costs for model restoration/retraining • Unauthorized access to AI environments (LLMjacking) • Privacy breaches involving AI systems	• Algorithmic decision errors • Misguidance causing client harm • Training-data misuse (negligence) • Media/IP liability for AI-generated outputs (where coverage available) • Bias/discrimination in automated decisions (where coverage available)	• Limited coverage for BI/PD if AI is incidental to product/service and not excluded (subject to standard terms)	• Defence costs for shareholder suits alleging mismanagement or misrepresentation of AI strategy • Coverage for governance failures under standard wrongful act definitions	• Defamation, libel, and slander from AI-generated output • Copyright/trademark infringement in outputs • Privacy violations in published AI content	• Model-error coverage for automated employment decisions (such as hiring) • Algorithmic decision errors and biased/discriminatory outcomes produced by AI-enabled tools, including defence against employment regulators (e.g., EEOC proceedings); damages/settlements where insuring clauses allow • Coverage for procedural violations tied to AI-assisted screening where policy language extends to wrongful employment practices with human oversight
Areas Commonly Excluded or Not Covered	• Criminal acts or fraudulent AI deployment • Prohibited AI practices under regulations • Model performance guarantees or warranties	• Intentional IP infringement • Performance guarantees for AI models • Model drift or underperformance without negligence • Certain regulatory penalties	• AI-generated personal & advertising injury • BI/PD from autonomous AI decisions • Broad AI exposures (excluded via emerging endorsements)	• Explicit AI operational failures • Model performance issues • Certain regulatory fines (insurability varies) • Fraudulent statements about AI capabilities	• Intentional harmful content • Training-data IP infringement (often excluded) • Algorithmic bias unless endorsed	• Intentional discrimination, violations of law, or prohibited practices; conduct exclusions still apply • Wage and Hour, Fair Labour Standards Act exposures (often excluded or tightly sub-limited) • Training-data IP disputes and non-employment privacy claims
Ambiguities	• Deepfake/social engineering losses (often sub-limited) • AI-driven outages without a security breach • Regulatory fines (jurisdiction-dependent)	• Coverage for AI outputs when there is no human oversight • Misrepresentation of AI capability (may trigger conduct exclusions) • IP disputes tied to training datasets	• Product liability for AI-driven errors • Defamation from AI-generated content • Attribution of fault between the AI vendor and the insured	• Oversight gaps vs. technical failures • Disclosure obligations under evolving AI regulations • Allocation of liability for third-party AI reliance	• Whether AI-generated content constitutes a “publication” under legacy wording • Coverage of synthetic media (deepfakes) used maliciously	• Allocation of liability between the employer and third-party AI vendors (who is responsible for tool performance vs. deployment/oversight) • Coverage where no human-in-the-loop is present or where documentation is weak



Affirmative Coverage

An emerging category of insurance offers dedicated coverage for AI model error risks. These new policies address ambiguities and potential gaps in traditional commercial insurance cover, particularly where probabilistic model behaviour is assessed through legacy negligence or "wrongful act" constructs. Variations exist, but at its core, affirmative coverage typically has an "all-risk" basis, specifically designed to cover liability arising out of model error, including scenarios that may not involve a cyber event or malicious actor.

One approach is to assess the "target model metric" as part of the underwriting process. The concept considers factors such as model accuracy levels and expected variance of outcomes to create a benchmark for each individual model. This provides a threshold below which model error is deemed to occur, a more objective measure than relying on a "wrongful act" definition, which can be inconsistently applied and lends itself to subjective claims determination.

Each model is underwritten on its individual merits, based on factors such as the industry, context of the outputs, any underlying foundation model, the version deployed and the use case. This allows for a bespoke approach to pricing, and an appetite to take on risk with a clear-sighted picture of the AI and its implications. This approach also enables clearer articulation of coverage intent and more disciplined exposure management at both the policy and portfolio level. Importantly, these approaches do not guarantee model performance, but define insurable thresholds for model failure events in a way that is allows underwriting to be auditable, governable, and scalable.

Examples of this type of affirmative AI coverage approach are already present in the market, including specialist AI liability solutions developed by underwriters such as Armilla AI. They focus on model-specific risk assessment and clearly defined triggers for AI failure events rather than extending traditional cyber or technology E&O wordings through endorsements.

Regulatory Coverage Expansion and the Role of Standards

"Regulation needs to catch up with innovation," observed Henry Paulson[7], the former US Treasury Secretary. He was referring to the financial trading markets, but his point was that regulation, by definition, lags behind the technology it is intended to create guardrails for. One example is the rapid rise of social media companies. Their explosive growth was evident well before their impact was fully understood, particularly on children's mental health, and the subsequent regulatory interventions. Similarly, the EU General Data Protection Regulation (2018) was fully implemented more than six years after it was initially proposed in a rapidly changing network and internet environment.

Sound regulation evolves over time and follows the path of technology adoption. But critically, effective regulation depends on laying out principles and providing guidance on how to achieve them. In this regard, the evolution of standards is equally important. Effective standards establish clear expectations of what "good" looks like and provide all parties – regulators, organisations, and insurers –with a common framework for measuring and mitigating risks. Standards enable underwriting and pricing at scale based on a common benchmark. Cyber insurance standards, such as ISO 27001 and the US National Institute of Standards and Technology (NIST) Cyber Security Framework, are used to set clear expectations for security posture, allowing underwriters to assess risk consistently against independent criteria across a portfolio.

Both regulations and standards relating to AI are still evolving. On the regulatory side, the EU AI Act is the most prominent development. Certain insurers have already developed coverage expansions for specific regulatory risks, such as those created by the EU AI Act and the Colorado AI Act. Both acts provide a risk-based approach to managing higher-risk AI systems, emphasizing transparency and accountability in model development. On the standards side, initiatives such as ISO 42001 and the NIST AI Risk Management Framework are maturing, though they still lack clear anchoring in regulation. The EU AI Act, for instance, has yet to specify harmonised standards that would allow a presumption of conformity. However, as both regulations and standards mature and converge, we expect coverage to evolve in line with them – addressing regulatory and compliance risks with greater precision.

Fines for breaches of prohibited AI practices under the EU AI Act can reach €35 million or 7% of annual revenue, with proportionately smaller fines for lesser breaches. Some of these fines are unlikely to be insurable, though associated costs – such as defence expenses, remediation, and crisis response – could be covered. As with any regulatory oversight, sound public policy reasons limit the insurability of fines and penalties. Coverage for regulatory exposure is typically offered via a sub-limit and may require supplemental applications. Pricing methodology remains somewhat opaque, given the limited examples of regulatory enforcement in this space.

Insurers are able to use regulations as a framework to develop underwriting approaches. These cover a range of AI-related issues that give insight into aspects of risk mitigation.

[7] https://home.treasury.gov/news/press-releases/hp897

Claim Scenario: AI Chatbot goes rogue

This example scenario illustrates errors that could easily happen in an AI system and how the risk could fail to trigger any existing commercial policies.

1. Company Profile

A $250 million outdoor equipment retailer, Camping Retail Corp, rolls out an AI-powered customer service agent using retrieval-augmented generation to pull from its product catalogue, warranty details, and return policies. The system handles about 5,000 customer interactions per day, with humans reviewing only escalated cases.

2. The Initial Error

A routine product catalogue update contains a script error that changed the warranty on high-end tents to "lifetime warranty" instead of the correct "two-year limited warranty." The mistake passed automated checks because the data format was valid, and the update went live without manual review.

3. Propagation and Discovery

Over the next several weeks, the AI agent responds to hundreds of customer inquiries about equipment warranties, consistently stating that the products carry lifetime warranties. The agent's responses are confident and specific, citing individual product references and warranty terms that appear in its context window. Several patterns emerge:

- Sales inquiries: Prospective customers purchase based on the lifetime warranty claim, paying premium prices compared to competitors. Business-to-business customers receive product specifications with incorrect warranty terms, which they incorporate into their own marketing materials.
- Warranty claims: Existing customers with tents outside the actual two-year warranty period submit claims, which the AI agent initially validates and processes.
- Return-period extensions: The agent tells customers they can return tents beyond the standard 30-day window based on misunderstood warranty language.

The AI agent's performance metrics during this period show excellent response times, high customer satisfaction scores, and resolution rates within normal parameters. No alerts are triggered because the agent is functioning exactly as designed; the problem lies in the source data, not the model itself.

The issue is discovered when the warranty team sees a spike in claims beyond the standard two-year limit. Investigators then find customer emails with screenshots of the AI agent promising lifetime warranties. A check of the knowledge base confirms the underlying data corruption.

4. Impact Assessment

The company's legal and operations teams assess the impact:

- **Contractual obligations:** Hundreds of customers have written confirmations from the AI agent stating lifetime warranty coverage. Legal counsel advises that these constitute binding commitments under contract law. Nearly 100 customers made purchases specifically citing these warranty terms.
- **Completed claims:** The company has already honoured dozens of warranty claims for products outside the actual warranty period, costing tens of thousands of dollars in parts and labour.
- **Reputational exposure:** Multiple customers have posted on social media and review sites about the "bait and switch" when they discovered the actual warranty terms.
- **B2B complications:** Multiple reseller partners have published marketing materials featuring the incorrect warranty terms and now face their own customer complaints.
- **Regulatory notification:** The state attorney general's office opens an inquiry into deceptive trade practices after receiving consumer complaints.

5. Resolution and Costs

The company chooses to honour commitments made by the AI agent while implementing new controls, resulting in the following example costs:

Direct warranty obligation costs	$450,000
Legal and regulatory response costs	$150,000
System audit and remediation	$45,000
Customer retention and goodwill programs	$80,000
Total estimated loss	$725,000

6. Insurance Implications

This scenario reveals critical gaps in traditional insurance coverage:

- **CGL:** The CGL policy excludes cyber-related losses. It may cover the "advertising injury" component, No products are defective so product liability coverage does not apply. An AI chatbot making incorrect statements will not typically constitute advertising. The contractual obligations created by the agent do not constitute bodily injury or property damage.
- **Cyber insurance:** This typically covers data breaches and related exposures, but it is not intended to cover AI agents providing incorrect information. There was no security failure or malicious actor – just an operational error leading to incorrect outputs.
- **Professional liability:** As an outdoor equipment retailer, they do not have professional liability insurance in place. This coverage is not in scope.

Evolving Systemic Risk

One phrase that causes the hairs to stand up on the back of insurance executives' necks is "systemic risk." The potential for events to impact multiple entities simultaneously or create cascading failures presents fundamentally different challenges when insuring AI risks compared to standard commercial lines. While many traditional policies address risks that typically manifest as individual losses, systemic AI vulnerabilities arise from structural characteristics that create inherent correlation across seemingly diverse portfolios.

Lessons can be drawn from other classes of insurance that share common sources of systemic risk. Cyber as a class of business has parallels, as the dependence on common technologies has grown. Similarly, E&O losses have the potential to accumulate within industry sectors based on macroeconomic factors. Examples include correlated risks such as fraud among the legal and accounting professions, which can increase during an economic downturn. In the world of property risk, catastrophe risk, such as earthquakes and hurricanes (e.g., Hurricane Andrew in 1992), transformed the way claims, underwriting, and capital were managed.

Traditional commercial insurance achieves portfolio diversification through geographic distribution, industry variation, and operational independence as a control against systemic risks. A cyber incident typically results from a specific organisation's security posture, and an E&O claim emerges from the delivery of particular professional services.

AI risk operates differently. Therefore, traditional systemic controls are not as effective. Organisations across different sectors may deploy functionally identical AI infrastructure, the same foundation models, similar architectures, and common data infrastructure specialised for AI. When a widely deployed model contains compromised training data or other unintended performance characteristics, failures can propagate simultaneously across multiple organisations regardless of geography, industry, or individual risk management practices. In addition, this can be exacerbated if the same underlying foundational model is used across different modalities.

AI systemic risks often stem from architectural characteristics inherent to the technology itself. The UK National Cyber Security Centre acknowledges that prompt injection "may simply be an inherent issue with LLM technology."[8] Similarly, the Open Worldwide Application Security Project (OWASP),[9] a nonprofit organisation that works to improve software security, notes that "as yet there are no surefire mitigations" relating to AI. A recent article by the Australian Signal Directorate and the Australian Cyber Security Centre claims approximately 25% of organisations have experienced AI data poisoning incidents,[10] while sophisticated attack techniques can achieve a 95% success rate in compromising widely used foundation models.[11]

AI prompts can be designed to specifically bypass safety considerations, a term known as AI jailbreaking. These techniques can become rapidly and widely known and simultaneously compromise systems across multiple organisations within days, regardless of an individual company's security maturity. This represents a different risk profile than traditional cyber exposures: not a failure of controls, but exploitation of fundamental model characteristics affecting all implementations.

Traditional commercial risks usually change gradually. Cyber threats shift as hackers research potential vulnerabilities and develop strategies; physical operations maintain relative stability, and property perils evolve in line with risks, such as climate change. Annual policy periods align reasonably well with risk evolution timeframes.

AI systems evolve at speed via rapid model updates, architectural changes, and deployment pattern shifts. A portfolio of AI risks that appears well-diversified at inception can develop a severe concentration of risk as organisations adopt similar newly released models or converge on common architectures. The pace of change in AI model scale is evidenced by the fact that training compute power is estimated to double every five months, datasets every eight months, and electricity power usage annually, according to the Stanford AI Index 2025.[12]

Implications for the Insurance Industry

Effective underwriting of AI risk requires a fundamentally different approach compared to traditional commercial lines. In addition to focusing on individual policyholder risk management practices, underwriters must evaluate portfolio-level exposure concentration through shared model dependencies, architectural vulnerability to coordinated attacks, and the capacity to detect failures before substantial liability accrues.

The challenge for the insurance industry is not whether AI will create systemic risk events, but when, and if underwriting practices can keep pace. AI risk concentration operates through different mechanisms than traditional commercial insurance. Geographic and industry diversification provide insufficient mitigation when multiple policyholders deploy functionally near-identical systems with common vulnerabilities. The tools that enable effective traditional commercial insurance portfolio management, performance metrics, risk maturity assessments, and loss history analyses become unreliable indicators when applied to AI systemic risk.

> “Effective underwriting of AI risk requires a fundamentally different approach compared to traditional commercial lines.”

> “There is a qualitatively different nature to AI technology, and it comes with novel risks that must be addressed to unlock its full potential.”

As the market develops, insurers must balance the significant opportunity AI insurance represents against fundamental uncertainty created by risk concentration mechanisms that differ structurally from other classes, such as cyber, E&O, and CGL risks. The question is not how traditional commercial insurance frameworks can be adapted for AI risk but whether entirely new approaches to portfolio management, loss correlation analysis, and catastrophic exposure modelling can be developed.

A Bright Future

Today's generative AI is only the first wave of a broader AI era. For those focused on addressing the risks it creates, the opportunities ahead are substantial. The existing commercial insurance product landscape has adapted in some areas, but it is not fit for purpose overall, given the speed of change.

The massive benefits that AI brings are clear, though the related risks are more opaque. Insuring these risks will be as important as the marine insurance offered to early exploratory sailing ships. Whenever there is ambiguity in insurance contracts, this can reduce the perceived value of the coverage provided. When it comes to AI, clarity is sought, both by policyholders and insurers. The sooner clarity is established for AI risks, the better.

There is a qualitatively different nature to AI technology, and it comes with novel risks that must be addressed to unlock its full potential.

[8] National Cyber Security Centre, "Prompt Injection"

[9] OWASP, "About the OWASP Foundation."

[10] Australian Signals Directorate, "Artificial Intelligence and Machine Learning"

[11] Pathade, "Red Teaming the Mind of the Machine"

[12] Stanford University. "Artificial Intelligence Index Report 2025."

Authors and Contacts

AUTHORS

Lockton Re

Oliver Brew ACII
Head of Cyber Centre of Excellence
Lockton Re
+44 (0)7384 248 268
oliver.brew@lockton.com

Eden Fall-Bailey
Cyber Account Executive,
Lockton Companies
eden.fallbailey@lockton.com

Armilla

Baiju Devani
CTO & Cofounder,
Armilla AI
baiju@armilla.ai

Alec Eyre
Head of AI Assessments,
Armilla AI
alec@armilla.ai

CONTACTS

Lockton Re

Matthew Silley FIA
Cyber Practice Leader, International
+44 (0)7391 387 699
matthew.silley@lockton.com

Jemima Hopper ACII
Cyber Reinsurance Broker
+44 (0)7350 436 652
jemima.hopper@lockton.com

Brian Lewis
US Cyber Practice Leader
+1 646 279 1940
brian.lewis@lockton.com

Caitlin Barnett
Cyber Reinsurance Broker
+1 929 675 9132
caitlin.barnett@lockton.com

Armilla

Karthik Ramakrishnan
CEO & Cofounder,
Armilla AI
karthik@armilla.ai

MEDIA CONTACTS

Lockton Re

Isabella Gaster
Lockton Re Global Head of Marketing
+44 (0)7795 400981
isabella.gaster@lockton.com

Elizabeth Miller Kroh
Lockton Re Head of Marketing, North America
+1 (445) 248 2228
elizabeth.kroh@lockton.com

Sources

1. St. George, Benjamin. 2023. "What Is the Turing Test?" SearchEnterpriseAI. April 2023. https://www.techtarget.com/searchenterpriseai/definition/Turing-test.

2. Wikipedia. "Artificial Intelligence." Last modified January 16, 2026. https://en.wikipedia.org/wiki/Artificial_intelligence.

3. Singh, Shubham. "ChatGPT Statistics for 2023: Comprehensive Facts and Data." Demand Sage. February 28, 2025. https://www.demandsage.com/chatgpt-statistics/.

4. Buolamwini, Joy, and Gebru, Timnit. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." Edited by Sorelle Friedler and Christo Wilson. Proceedings of Machine Learning Research 81, no. 1 (2018): 1–15. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf.

5. Swiss Re Institute. "How Deepfakes, Disinformation and AI Amplify Insurance Fraud." 2025. https://www.swissre.com/Institute/Research/Sonar/Sonar2025/How-Deepfakes-Disinformation-Ai-Amplify-Insurance-Fraud.html. Swiss Re Institute. June 12, 2025. https://www.swissre.com/institute/research/sonar/sonar2025/how-deepfakes-disinformation-ai-amplify-insurance-fraud.html.

6. "Automated Employment Decision Tools (AEDT) | DCWP." n.d. Www.nyc.gov. https://www.nyc.gov/site/dca/about/automated-employment-decision-tools.page.

7. "Remarks by Secretary Henry M. Paulson, Jr. On Blueprint for Regulatory Reform." 2008. U.S. Department of the Treasury. March 31, 2008. https://home.treasury.gov/news/press-releases/hp897.

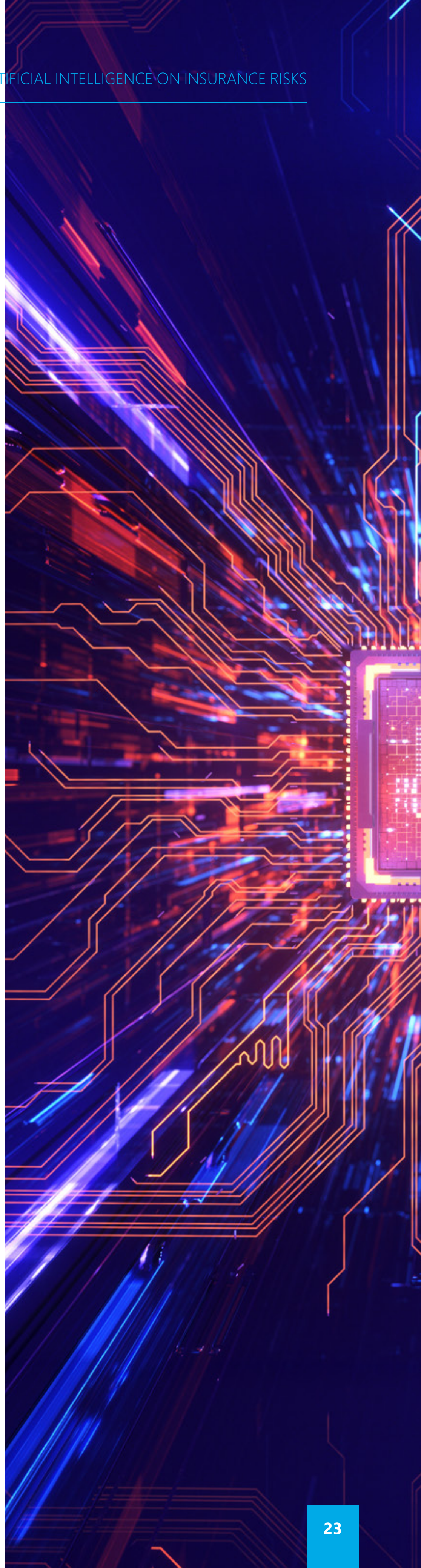
8. National Cyber Security Centre. "Prompt Injection Is Not SQL Injection (It May Be Worse)." 2025. https://www.ncsc.gov.uk/blog-post/prompt-injection-is-not-sql-injection.

9. OWASP. Accessed January 2026. "About the OWASP Foundation." Owasp.org. https://owasp.org/about/.

10. Australian Signals Directorate. "Artificial Intelligence and Machine Learning: Supply Chain Risks and Mitigations." October 16, 2025. https://www.cyber.gov.au/business-government/secure-design/artificial-intelligence/artificial-intelligence-and-machine-learning-supply-chain-risks-and-mitigations.

11. Pathade, Chetan. "Red Teaming the Mind of the Machine: A Systematic Evaluation of Prompt Injection and Jailbreak Vulnerabilities in LLMs." arXiv. May 7, 2025. https://arxiv.org/html/2505.04806v1.

12. Stanford University 2025. "Artificial Intelligence Index Report 2025.". https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf.



www.armilla.ai

1179 King St W Suite 112, Toronto, ON M6K 3C5, Canada

This publication is provided for general informational purposes only and does not constitute insurance, legal, tax, accounting, or other professional advice. Descriptions are general in nature only; please refer to the specific terms, conditions, exclusions, and eligibility requirements of any insurance policies offered or purchased.

Insurance services are provided by Armilla Insurance Services, Inc., 200 Continental Drive, Suite 401, Newark, Delaware 19713, United States of America. Insurance products are subject to underwriting approval and regulatory requirements and may not be available in all jurisdictions. Pricing depends on a variety of factors, including policyholder location. Recommendations, examples, or illustrations are for informational purposes only and do not guarantee coverage, pricing, or outcomes.

REINSURANCE

www.locktonre.com

261 Fifth Avenue, New York • NY 10016

The St. Botolph Building, 138 Houndsditch • London EC3A 7AG

Lockton Re is a trading name and logo of various Lockton reinsurance broking entities and divisions globally and any services provided to clients by Lockton Re may be through one or more of Lockton's regulated businesses.

Our regulated entities are: Lockton Re, LLC in the USA, Lockton Re, LLC, 261 Fifth Avenue, 10th Floor, New York, NY 10016, in California, Lockton Re Intermediary Insurance Services, LLC (License number 0G76373); Lockton Re LLP in the UK Registered in England & Wales at The St. Botolph Building, 138 Houndsditch, London, EC3A 7AG. Company number OC428915 and authorized and regulated by the Financial Conduct Authority with FRN 921278; Lockton Re (Bermuda) Limited in Bermuda (Registration number 56292) Seon Place, 141 Front Street, Hamilton HM19, Bermuda; Lockton Re (Germany) GmbH, Ludwigpalais, Ludwigstraße 8, 80539, Munich, registered with the Commercial Register of the Munich Local Court under HRB 275178, authorized as an insurance broker pursuant to Section 34d para. 1 GewO and regulated by the Chamber of Commerce (IHK), Munich and Upper Bavaria with Company Registration Number: D-I5DQ-33IIL-38; and Lockton Re is a trading name of Lockton (MENA) Limited. Lockton (MENA) Limited is registered in the Dubai International Financial Centre (DIFC), Gate Village Building 06, DIFC PO Box 506794 Dubai, UAE commercial license number 0970, and is regulated by the DFSA, the independent financial services regulator for the DIFC; Lockton Re LAC Series of Lockton Specialties, LLC, (Latin America & Caribbean) registered under license number 8331631 in the Missouri Department of Commerce and Insurance and under license number W697048 in the Florida Department of Financial Services, with its business address at 1221 Brickell Avenue, Suite 1500, Miami, FL 33131, USA; Lockton Re Brasil Corretora de Resseguros Ltda., (Brazil) registered under permit number 7,610 at SUSEP - Superintendência de Seguros Privados, with its business address at Avenida das Nações Unidas, 14171, cj. 1401, São Paulo, SP, Brazil; Lockton Chile Corredores de Reaseguros SPA., (Chile) registered under license number C-282 at CMF - Comisión para el Mercado Financiero, with its business address at 90 Orinoco, office 2001-A, Las Condes, Santiago, RM, Chile; Lockton México Intermediario de Reaseguro, S.A. de C.V., (Mexico) authorized by Comisión Nacional de Seguros y Fianzas, with its business address at Av. Santa Fe 481, Piso 20, Colonia Cruz Manca, Cuajimalpa. C.P. 05349, Mexico; Lockton Corredores de Reaseguros Ltda., (Colombia) registered under Resolution number 1801 at Superintendencia Financiera de Colombia, with its business address at Carrera 11 No 86 - 60, Bogotá D.C., Cundinamarca, Colombia.

Lockton Re provides this publication for general informational purposes only. This publication and any recommendations, analysis, or advice provided by Lockton Re are not intended to be taken as advice regarding any individual situation and should not be relied upon as such. It is intended only to highlight general issues that may be of interest in relation to the subject matter and does not necessarily deal with every important topic nor cover every aspect of the topics with which it deals. The information and opinions contained in this publication may change without notice at any time. The information contained herein is based on sources we believe reliable, but we make no representation or warranty as to its accuracy. Lockton Re shall have no obligation to update this publication and shall have no liability to you or any other party arising out of this publication or any matter contained herein. Any statements concerning actuarial, tax, accounting, or legal matters are based solely on our experience as reinsurance brokers and risk consultants and are not to be relied upon as actuarial, tax, accounting, or legal advice, for which you should consult your applicable professional advisors. Any modeling, analytics, or projections are subject to inherent uncertainty, and the information contained herein could be materially affected if any underlying assumptions, conditions, information, or factors are inaccurate or incomplete or should change. This publication is not an offer to sell, or a solicitation of any offer to buy any financial instrument or reinsurance product. If you intend to take any action or make any decision on the basis of the content of this publication, you should seek specific professional advice and verify its content.

Lockton Re specifically disclaims any express or implied warranty, including but not limited to implied warranties of satisfactory quality or fitness for a particular purpose, with regard to the content of this publication. Lockton Re shall not be liable for any loss or damage (whether direct, indirect, special, incidental, consequential or otherwise) arising from or related to any use of the contents of this publication.