

# ACTUARIAL INTELLIGENCE BULLETIN



March 2026

## Table of Contents

AI Insights for Actuaries: A Virtual Symposium.....	2
SOA Research Institute AI Coverage in 2026: What to Expect.....	3
SOA Artificial Intelligence and Actuaries Survey 2026.....	5
Health Insurance in the Age of Big Data and AI.....	5
Why Generative AI Makes APL-to-Python Modernization Realistic.....	9
When AI Meets Insurance: Why Underwriting is at the Center of Industry Change .....	10
Using AI to Do Research About AI .....	12
Can AI Implement Actuarial Model Changes? A Practical Experiment .....	14
How One Experiment with Copilot Changed My Entire Workflow .....	17
Understanding AI Risk Beyond the Model: A Conversation with Karen Bennet .....	18
Why AI Can Systematically Amplify Risk — and What Actuaries Can Do About It .....	20
Can I Trust My AI Output? Prompt Design, Variance, and Validation of LLM-driven Insurance Systems .....	23
SOA Expert Panel Report: The AI-Ready Actuary.....	26
AI Will Not Replace Actuaries, but Actuaries Who Adapt Will Replace Those Who Don't.....	28
The Pilot and the Plane: Actuary-in-the-Model.....	29
Editorial Committee .....	31
About the Society of Actuaries Research Institute .....	32

Welcome to the March 2026 edition of the SOA Research Institute AI Bulletin! This bulletin serves as a platform for sharing knowledge and fostering collaboration around artificial intelligence within the actuarial community. Explore articles on strategic initiatives, practical tips, and research advancements, all aimed at empowering actuaries to leverage AI responsibly and effectively.

### Caveat and Disclaimer

The opinions expressed and conclusions reached by the authors are their own and do not represent any official position or opinion of the Society of Actuaries Research Institute or the Society of Actuaries or its members. The Society of Actuaries Research Institute makes no representation or warranty to the accuracy of the information.

## AI Insights for Actuaries: A Virtual Symposium

JON FORESTER, ASA, MAAA

### Virtual Meeting: May 13–14

Artificial intelligence is no longer a future concept for the actuarial profession—it is a present-day capability. As AI reshapes how data is analyzed, risks are modeled, and insights are communicated, actuaries who understand how to harness these tools will be best positioned to lead.

### Learn by Doing: Live, Hands-On AI Demonstrations

A defining feature of the May 13–14 meeting is its focus on **live, hands-on demonstrations**, allowing attendees to see how AI tools work in practice and how they can be applied to real actuarial challenges. Featured sessions include:

- **Sherry Chan and Joseph Bonezelli** address *answers to the most common AI questions*, cutting through complexity and hype
- **Eric Milner** demonstrating how AI can be leveraged in the *Monté Carlo simulation*
- **Shirly Wang** exploring *how AI agents empower human actuaries*
- **Michael Niemerg** presenting approaches for *AI-driven data analysis*
- **Eric Wolfe and Sian Walker** covering *Gen-AI documentation*, with a focus on quality, consistency, and governance
- **And more!**

These sessions are designed to move actuaries from awareness to confident application.

### Ethics, Bias, and Professionalism

Responsible AI use remains central to the actuarial profession. The virtual May meeting will feature dedicated sessions on **bias and professionalism**, reinforcing the importance of ethical judgment, transparency, and professional standards when applying AI in actuarial work. These sessions will fulfill your 1 hour of bias CE credit and count toward your professionalism CE credit requirement.

### Management Track: Leading AI Adoption

For actuaries in leadership and management roles, the **Management Track** focuses on strategy, governance, and organizational readiness:

- **Dave Ingram** on cultivating an effective *AI culture*
- **Igor Nikitin** discussing *AI for C-level decision makers*

### Become the AI-ready actuary of the future Free Webcast: April 30

The third annual **AI Insights for Actuaries** event offers multiple ways to engage, beginning with a **free webcast on April 30**, followed by the full meeting, live on **May 13–14**.

To help actuaries begin—or deepen—their AI journey, a **free webcast on April 30** will be offered to all interested participants. This webcast is open to everyone and will deliver valuable AI insights relevant to actuarial work today.

For those attending the main conference, the webcast provides an excellent foundation and shared context ahead of the live sessions. For those unable to attend the meeting, it offers an accessible opportunity to learn about AI trends, applications, and considerations shaping the profession.

- **Mark Sayre** examining *legal and regulatory considerations*, bringing a unique perspective as both an actuary and a lawyer

This track equips leaders to guide AI adoption thoughtfully and responsibly.

### Practical Track: Applied AI Skills

The **Practical Track** is designed for actuaries looking to build hands-on skills they can use immediately:

- **Arthur da Silva** on *AI agents*
- **Simon Hatzesberger** on *LLM-powered knowledge graphs*
- **Ian Pietz** on *narrow AI models* and targeted solutions

These sessions emphasize practical tools and real-world use cases.

Stay tuned to find the full agenda for each track, coming soon!

### Bonus Learning with PD Edge+ Video Content

Attendees will also receive access to **five bonus on-demand videos from PD Edge+**, extending learning beyond the live event and allowing participants to revisit key topics at their own pace.

We are looking forward to another engaging AI Insights symposium. I will see you there.

*Jon Forster, ASA, MAAA is Professional Development Learning & Content Design Director at the SOA.*



## ACTUARIAL INTELLIGENCE BULLETIN

## SOA Research Institute AI Coverage in 2026: What to Expect

**DALE HALL, FSA, CERA, CFA, MAAA**

AI will remain a major focus for the SOA in 2026, and the SOA Research Institute plans to keep its AI work practical, standards-aware, and directly usable by actuaries in day-to-day roles. The through-line is simple: AI is moving fast, so the Research Institute's work will emphasize what can be applied now (use cases and methods), what must be controlled (governance and risk management), and what must be monitored over time (skills, adoption patterns, and emerging gaps).

### Staying Close to the Standards Conversation

A cornerstone of the 2026 plan is the Research Institute's continued active involvement in the U.S. AI Consortium run through the National Institute of Standards and Technology (NIST). The Research Institute has participated since early 2024 and intends to remain engaged as AISIC develops standards, guidelines, and testing environments for safe, secure, and trustworthy AI under the NIST.

For actuaries, this matters because "trustworthy AI" is not a slogan—it must translate into defensible practices: testing, documentation, controls, and ongoing monitoring. The AI Consortium's areas of emphasis (including red-teaming, risk management, synthetic content, and AI workforce skills) align closely with what insurance and retirement organizations increasingly need when AI is deployed in regulated, high-stakes environments.

## A Broad Research Pipeline

The Research Institute currently has a large portfolio of AI-related research projects underway (16 in progress) with releases expected in 2026. Rather than concentrating on a single practice area, the work spans insurance and retirement topics and reflects a deliberate mix of foundational guidance and applied demonstrations.

The portfolio is best understood through its themes:

### Governance and Ethics for Modern AI (Especially GenAI)

Several projects focus on what “good governance” should look like when models are generative, outputs can vary with prompts, and traditional validation approaches don’t always map cleanly. Expect emphasis on practical governance, such as how to design controls, documentation, testing, and accountability that are workable in our actuarial applications.

### Fairness, Bias, and Data Representativeness

The pipeline includes work on fairness and bias in insurance contexts and on how representation issues in real-world datasets can affect conclusions and outcomes. The aim is to move beyond general principles and toward methods that actuaries can use to evaluate issues, compare mitigation strategies, and communicate limitations.

### Applied AI in Actuarial Workflows

The Research Institute’s research agenda includes work tied to operational use cases—areas where organizations are already experimenting and where actuarial oversight is essential. This includes topics connected to underwriting and pricing impacts, management reporting, and reserving techniques, with attention to measurable value and measurable risk. In addition, we are partnering with several universities that are already studying the next generation of AI applications for actuaries.

### Retirement Applications

Another thread addresses AI’s influence on retirement professionals and retiree decision-making, including how AI can support planning and where human domain expertise remains critical. This is a space where trust, suitability, and explainability are especially important.

### Workforce and Skill Implications

A set of projects is aimed at how AI changes technical skill requirements for actuaries and how automation may reshape employment and investment dynamics. The objective is not hype, but clarity: what skills are becoming table stakes, what gaps are emerging, and what that implies for actuarial teams.

### More AI projects Will Launch During 2026

The Research Institute does not view the current 2026 pipeline as “the list.” It expects to initiate additional AI research projects throughout 2026. That approach is intentional: AI capabilities, regulatory expectations, and practical risks are evolving too quickly for a static research plan. A rolling portfolio is how the Research Institute intends to stay relevant as new issues surface.

### Tracking Change: A Refreshed AI Member Survey

Because adoption patterns shift quickly, the Research Institute also plans to repeat an AI member survey in 2026 (building on the 2025 survey). The goal is to measure change over time—capturing what actuaries are using AI for, how practices are maturing, and where gaps or opportunities remain. In other words, it is meant to keep the research agenda grounded in what actuaries are encountering, not just what is theoretically interesting.

## What This Means in Practice

In 2026, the Research Institute's AI coverage is positioned to do three things:

1. **Connect actuarial practice to the broader safety-and-standards ecosystem** through sustained AISIC involvement.
2. **Deliver research that is both practical and controls-oriented**—focused on governance, fairness, validation, and real workflows, not abstract commentary.
3. **Keep the agenda current** through additional new projects and a survey-based feedback loop that identifies where the profession is progressing and where support is still needed.

Bottom line: the Research Institute's 2026 AI plan is built around helping actuaries use AI responsibly and effectively—while keeping rigor, controls, and real-world measurability at the center.

*R. Dale Hall, FSA, MAAA, CERA, CFA is Managing Director of Research at the Society of Actuaries Research Institute.*



## ACTUARIAL INTELLIGENCE BULLETIN

## SOA Artificial Intelligence and Actuaries Survey 2026

The Society of Actuaries Research Institute's Actuarial Innovation and Technology (AIT) Research Program invites you to participate in an important survey on the use of Artificial Intelligence (AI) in actuarial work, including the growing impact of generative AI tools. This is the 2026 update to the survey that was conducted in 2025.

2026 Survey on the use of Artificial Intelligence (AI) in actuarial work.

[SURVEY LINK](#)

Whether you are actively using AI or just beginning to explore its potential, your perspective is vital to guiding future strategic initiatives and ensuring that the actuarial profession remains at the forefront of innovation.

The survey should take less than 10 minutes to complete, and all responses will be kept confidential.



## ACTUARIAL INTELLIGENCE BULLETIN

## Health Insurance in the Age of Big Data and AI

**IAN DUNCAN PHD FSA MAAA**

There has been considerable innovation in healthcare, and this is accelerating. Components of the innovation (but just a few - there are many others) are Technological advances, Big Data/AI/Machine learning. But here is the

This article is based on the inaugural address for the Duncan Laboratory for Insurance Data Science that Duncan gave at Heriot-Watt University's Panmure House in Edinburgh. Panmure House was the home of Adam Smith, author of *The Wealth of Nations*, so the talk focuses on economics.

conundrum: with all this innovation, technology and AI, why do health care costs continue to escalate at a faster rate than the overall inflation rate?

And what can actuaries contribute to reducing this problem?

### Some Background on U.S. Health Insurance

Uniquely expensive (average cost > \$10,000/annually per individual). By contrast, the median income in the US is \$61,000. Two main payers (~ 50/50): Government and Employers. Healthcare is a Third-party payment system. Health Insurance differs from other types of insurance:

- Driven by the tax code.
- Yearly renewable policies.
- Very high claim rates; (and highly-skewed amounts).
- Premiums need to fund current year's claims.
- High membership churn/turnover.
- Limited consumer choice/product innovation.

According to Milliman, health insurance costs grew at an annual compound rate of 5.6% between 2005-2025, compared with a CPI increase of 2.3%.

### The Role of Technology

Technology has improved the quality and reduced the cost of most services in my lifetime; examples include the supermarket, the ATM, the smartphone, Uber rides...the list is long. What do disruption and transformation of these industries have in common?

**ANSWER:** Through data and technology, activity is transferred to **consumers**, reducing **cost**, increasing **efficiency** and **consumer satisfaction**.

The consumer trades some additional labor for reduced cost, increased choice and satisfaction. While these services have become more efficient, healthcare has not.

While medicine has become more technologically-intensive, it is also a very human-resource-intensive business.

- You still have face-to-face (one-on-one) visits to the doctor.
- Aside from their technology, hospitals are still organized as they were by Florence Nightingale.

Other industries have used technology to increase consumer power and choice, and to decrease cost by shifting work to consumers. Why not medicine?

Innovation in healthcare has had a mixed record. Large amounts of capital have been devoted to it. Between 2019-23 digital health funding in the US totaled \$95 billion; globally the amount was \$142 bn. Half of healthcare investment in 2025 went to AI. While there have been some successes all this investment hasn't reduced cost, and ultimately someone will have to pay for investment in AI. There have been some spectacular failures in health tech as well (Google Health; IBM Watson; Theranos; Walmart Health; Haven Health....).

### My Hypothesis

In part it is because of the way medicine is financed. Either:

- By large insurance companies, in insurance systems (such as the U.S.) or
- By national health systems.

In both cases, decisions are made **for** the consumer, not **by** the consumer. The result is that the traditional market mechanism – price signals – doesn’t work. Consumers lack incentives to manage their health and their spending. (Cynics refer to this as the “all-you-can-eat-buffet” of American healthcare.)What is needed?

In my opinion, without these, no technology or AI solution will be successful:

- Payment Reform
- Data Analytics
- Behavioral Economics

**Payment Reform:** In the US we are experimenting with many different models (Pay-for-performance; gainsharing; capitation; bundled payments; ACOs); they tend to have one thing in common: they transfer risk from the payer to the provider of healthcare. But there is an important component missing in these models: the third player, the patient, does not bear a share of the risk.

**Data Analytics:** Data analytics/Predictive Analytics (and now AI) in business helps consumers make better choices, consistent with their long-term goals (cheaper, better products; tailored choices; faster access). Predictive analytics is now standard in healthcare and there are increasing numbers of AI models. A lot of ML/PA/AI innovation has reduced the labor component of healthcare services (both quantity and intensity).

- The conundrum is: why has this innovation not reduced the cost of services to insurers and consumers?
- Who has gained (captured the surplus) from the reduction in labor cost of services?

For example: how long will it be before the Primary Care physician is replaced by AI? The U.S. government predicts a shortage of 187,000 physicians by 2037. Many PCP tasks are ripe for replacement by AI: Diagnosis; Tests; Prescribing; Monitoring chronic patients. Already, the state of Utah allows renewal of prescriptions by AI without a doctor involved. But there are significant barriers to acceptance of the AI doctor: Will patients be willing to accept a bot as their physician if surplus is captured by physicians and payers? Incentives will need to change to change patient behavior.

**Behavioral Economics:** Behavioral economists rely on psychology rather than conventional economic principles to explain human behavior. Actuaries tend to focus on economic motivators for good reason: as the RAND experiment showed. The RAND Health Insurance Experiment (now almost 40 years ago!) addressed two key questions in health care financing:

- How much more medical care will people use if it is provided free of charge?
- What are the consequences for their health?

The RAND results showed that a 10% increase in cost-sharing would result in about a 2% reduction in spending. However, there were consequences as “good” utilization was reduced, along with “bad.” A less-blunt instrument is required.

The insurance model is a mystery to most consumers. It will have to change to support consumer healthcare behavior. There is a major role for AI in the future providing reliable information and advice for consumers that wish to manage their **healthcare budgets**.

## A Design for the Future

The current third-party payment system needs to change to include the consumer. We can achieve this with Individual Risk-adjusted Vouchers for each eligible participant, providing a basic budget for the patient based on medical conditions.

Patients will have a choice:

- They may contract with an insurer, as they do today – the insurer provides complete services. OR
- The patient may assemble their own portfolio of providers:
  - Inpatient/ER – contract with a neighborhood hospital
  - Urgent care
  - PCP/General Practitioner
  - Specialty care according to need (diabetes; cardiology etc.)
  - Other services (drugs; devices etc.) OR
- Provider Groups may offer bundles of different services.
- And Analytics Companies provide useful data/advice/decision support for Consumer Choice.

The key is **risk adjustment** of the underlying national average cost of care.

## The Healthcare Actuary of the Future

In the US, dynamics of Demand, Supply and Cost are leading to increasing demands from consumers for Government intervention in the market. Traditional economic tools (price signals; competition) don't seem to work in the healthcare market.

While there has been considerable innovation in healthcare these developments have had limited impact on rising costs. AI may help to make healthcare delivery more efficient but self-care and lack of consumer awareness of the financial implications of their choices remain the biggest challenge.

Actuaries combine the three components necessary for successful change:

- Payment reform.
- Data analytics/AI; and
- Behavioral Economics.

Actuaries understand the business and risk context in which medicine and healthcare operate and are experts at risk and risk adjustment. This will require acquisition of new skills, particularly in product design, pricing and consumer behavior. But no profession is better-suited to take advantage of the changes necessary in the future.

To do so, however, requires a new kind of health actuary with upgraded statistical and clinical knowledge.

Fortunately, many actuarial organizations are offering educational opportunities in these areas.

*Ian Duncan PhD FSA FIA FCIA FCA CSPA MAAA is President, DF Holdings Inc. and President-elect, Society of Actuaries 2025-6.*



**ACTUARIAL  
INTELLIGENCE  
BULLETIN**

## Why Generative AI Makes APL-to-Python Modernization Realistic

PRABHDEEP SINGH, FSA, MAAA, CERA, PMP

For decades, the programming language APL has played a central role in life actuarial modeling. Its expressive power and compact notation enabled sophisticated pricing, valuation, and projection systems long before modern actuarial software ecosystems existed. Many insurers still rely on actively maintained APL models that embody years—sometimes decades—of institutional knowledge.

Until recently, however, the idea of modernizing those environments into a language like Python often felt impractical. The technical risk, the reconciliation effort, and the scarcity of proven translation approaches made such projects difficult to justify. That calculus is now changing. Advances in generative artificial intelligence have made APL-to-Python modernization a realistic possibility for the first time.

### Why the Pressure to Modernize Has Been Building

Until a few years ago, automated code translation was largely limited to rule-based tools and narrow transpilers (a program that translates source code written in one programming language into source code written in another language). These approaches struggled with real-world codebases, especially those built around domain-specific idioms and implicit assumptions. Recent research suggests that this landscape has shifted materially.

Human judgment remains essential, but the cost and risk profile of experimentation has dropped significantly.

One strand of work focuses on large-scale translation efforts in technical computing domains, such as converting Fortran-based high-performance computing code into C++. These studies demonstrate that transformer-based models can preserve structure and intent across languages when trained and evaluated carefully, particularly when human review and domain knowledge are incorporated into the process.

A second, broader body of research surveys the state of generative AI for code translation more generally. A recent systematic mapping study by Rgaguena, Chlioui, and Radgui reviews more than 50 papers published between 2020 and 2025<sup>1</sup>, documenting rapid progress in large language models applied to code migration.

Together, these strands suggest an important shift: translation is no longer confined to toy examples or syntax-level rewrites. Instead, AI-assisted tools are increasingly capable of supporting real modernization efforts—provided they are used within a disciplined process.

### Why APL → Python Is Now Feasible

What has changed is not that these challenges have disappeared, but that the tooling landscape has expanded. Generative AI can now assist with:

- Initial code translation and refactoring
- Documentation of legacy logic
- Test generation and reconciliation support

---

<sup>1</sup> Rgaguena, A., Chlioui, I., & Radgui, M. (2025). Generative AI for code translation: A systematic mapping study. *Engineering Proceedings*, 112(1), 33. <https://doi.org/10.3390/engproc2025112033>

- Incremental redevelopment of model components

Python’s compatibility with these tools makes it a natural target environment. Instead of a wholesale rewrite, teams can use a hybrid approach: mechanically translating some components to preserve structure, while redesigning others in a more idiomatic, modular Python style.

Human judgment remains essential, but the cost and risk profile of experimentation has dropped significantly.

### Practical Takeaways for Actuarial Teams

For actuarial leaders and practitioners considering modernization, several practical points emerge:

1. Generative AI has lowered the barrier to entry. APL-to-Python modernization is no longer speculative; it is technically feasible with today’s tools.
2. Translation is only part of the problem. Sustainable modernization requires attention to education, governance, testing, and ongoing model development.
3. Push-button tools have a place—but also limits. Automated translation can accelerate progress, but actuarial review and domain expertise remain indispensable.
4. Python enables broader integration. Modern languages connect actuarial teams more naturally to analytics, data science, and AI-assisted workflows
5. Incremental approaches are viable. Modernization need not be all-or-nothing; hybrid strategies can reduce disruption while building internal capability.

The combination of long-standing organizational pressures and recent advances in generative AI marks a genuine inflection point. For the first time, modernizing APL-based actuarial models into Python is not just desirable—it is realistically within reach.

Disclosure: At least some parts of this article were written with the assistance of Generative AI, ChatGPT 5.2

*Prabhdeep Singh, FSA, CERA, MAAA, PMP Independent Consulting Actuary, Singh Actuarial Consulting, LLC.*



## When AI Meets Insurance: Why Underwriting is at the Center of Industry Change

**MARTY ARNOLD, FCAS, CERA**

Start Imagine two industries at very different ends of the spectrum.

One is slow to change—mature in its processes, highly regulated, and sitting on mountains of data it hasn’t fully learned to unlock. Employees in this industry build their work around long-standing relationships, often meeting in person, doing their best to use data that can be difficult to interpret, and making decisions informed by experience and judgment.

Now consider the other industry. It moves fast, operates with lighter regulation, and knows how to digest enormous volumes of data at speed. Relationships form quickly and often virtually. Processes are sleek, constantly evolving, and wherever possible, work is automated.

Welcome to the intersection of AI meeting insurance. One side was going to change—and it wasn’t going to be AI.

This merger of opposites is forcing change at a rapid pace, and underwriters find themselves at the center of a significant metamorphosis. For executives and strategists, this moment represents both opportunity and risk. As AI matures and stabilizes, carriers are advancing solutions that increase operational efficiency, automatically triage submissions, streamline renewals, price risk more accurately in real time, and sharpen risk precision. Collectively, these improvements reduce friction across the underwriting lifecycle.

Welcome to the intersection of AI meeting insurance. One side was going to change—and it wasn’t going to be AI.

As a result, underwriting skills are evolving. The value is no longer in doing all the work, but in serving as a *human-in-the-loop* during the broader “agentic” shift. This evolution demands stronger data literacy. As machines take on more routine tasks, underwriters must adapt—becoming informed consumers of data and thoughtful evaluators of AI-generated insights. Examples of AI-enabled continuous risk monitoring include:

- Tenant Occupancy Alerts – AI monitors real-time data to alert underwriters if a high-risk tenant moves into a property, which may require policy reclassification
- Geospatial Risk Shifts – Satellite imagery analyzed to detect vegetation density changes that increase a property’s susceptibility to wildfire
- Operational Drift - In Workers’ Compensation, AI compares payroll data with website updates to detect if a company has pivoted to doing higher risk work.

That transition, however, is not without resistance. Common concerns include:

- **Algorithmic bias**, with concerns that AI trained on historical data might be skewed by inadequate credibility or relevance of past data sets when projecting future exposures.
- **Data governance** and privacy controls making the underwriters responsible for keeping more data safe.
- **Understanding the AI results** and being able to explain decisions when some of the inputs into those decisions are “black box” algorithms.

Responding effectively requires intention and balance. Leading organizations are focusing on several key strategies:

- **Upskill and mentorship** between generations – teaming up generations helps to bridge the gap for carriers that struggle to transfer decades of institutional knowledge to a new generation that is tech-savvy and lacks deep risk assessment experience.
- **AI-Free skills** – organizations are starting to evaluate human-critical traits like complex problem-solving, empathy, and emotional intelligence that AI cannot replicate, ensuring employees have the right skills to drive the organization forward.
- **Recruitment** – organizations need to continue to engage the next generation, drawing new talent into an industry that has historically been seen as reluctant to change. Partnering with on-campus organizations like Gamma Iota Sigma and new university risk management programs to build the talent pipeline.
- **Retention** – finding ways to retain those underwriters that are versed in both AI and the traditional risk assessment skills.
- **Having a vision and orchestrating the change** – ensuring that the big picture is always top of mind for the total organization with incremental changes that move the right changes at the right times.

Artificial intelligence is moving rapidly from pilot programs into core insurance operations, with regulators and industry bodies highlighting its expanding role in underwriting, pricing, claims, and risk selection. What was once an AI-resistant industry is becoming increasingly AI-ready.

The future of underwriting will belong to professionals who are fluent in data and technology—yet still grounded in judgment, relationships, empathy, and the human insight required to partner effectively with brokers and customers.

*Marty Arnold, FCAS, CERA is Chief Underwriting Officer at Amerisure Insurance.*



## ACTUARIAL INTELLIGENCE BULLETIN

### Using AI to Do Research About AI

**MITCH STEPHENSON, FSA**

At a recent conference, the presenters asked the audience of more than three hundred — mostly actuaries — what they used artificial intelligence for at work. The top three responses were: 1) research, 2) summarization, and 3) coding. The topic of this article will be using AI to do research.

There are obvious reasons to do research with AI. One is to search for information. Most browsers include an AI generated response to accompany internet searches. A second is to identify sources applicable to a specific topic. A third is to compare and provide insight on a topic across sources. This article will explore the uses of identifying sources and comparing content across sources.

In May 2025, I published [From Zero to AI Governance](#) in the Actuarial Intelligence Bulletin.<sup>2</sup> It includes a table of five ethical principles associated with the use of AI, requirements to accompany the principles, and best practices to help meet those requirements. With new publications, specifically the [Three New Papers to Support Responsible AI in Actuarial Practice](#) released by the International Association of Actuaries (IAA),<sup>3</sup> I want to use AI to do research, and to recommend updates, to this table. I will use two large language models (LLMs), Claude Sonnet 4.5, and Chat GPT-5-MINI, for this exercise. Below are the prompts, results, and conclusions for this exercise of using AI to do research about AI.

**First prompt:** List the key references and cite specific articles, publications, and papers from actuarial sources on the risk, controls, and governance of artificial intelligence in actuarial work.

- **Result:** Claude lists twenty-three sources. It groups them by professional organization. This includes the Society of Actuaries (SOA), IAA, the Casualty Actuarial Society (CAS), and National Association of Insurance Commissioners (NAIC). Chat GPT groups its response by organization but describes types of sources instead of listing specific references. References by Claude that I expected to see are the IAA's [Artificial Intelligence](#)

---

<sup>2</sup> Mitchell Stephenson, "Zero to AI Governance: Establishing a Principles-Based Framework," in *Actuarial Intelligence Bulletin* (May 2025), 7–10, Society of Actuaries Research Institute, <https://www.soa.org/resources/research-reports/2025/research-ai-bulletin/>.

<sup>3</sup> International Actuarial Association, "IAA Releases Three New Papers to Support Responsible AI in Actuarial Practice," news release, December 3, 2025, <https://actuaries.org/news-post/iaa-releases-three-new-papers-to-support-responsible-ai-in-actuarial-practice/>.

[Governance Framework for General Actuarial Practice](#),<sup>4</sup> [SOA’s AI Risk Management Frameworks: An Expert Panel Discussion](#),<sup>5</sup> the CAS [Machine Learning in Insurance](#),<sup>6</sup> and the NAIC [Model Bulletin on the Use of Artificial Intelligence Systems by Insurers](#).<sup>7</sup> A notable omission is the [American Academy of Actuaries \(AAA\) Actuarial Professionalism Considerations for Generative AI](#).<sup>8</sup>

- **Observation:** Claude figured out from the prompt that I wanted each source listed with a link, but Chat GPT did not. An updated prompt to Chat GPT that gets this result is: “List, and provide links to, articles, publications, and papers released by global actuarial organizations on the use of artificial intelligence in actuarial work.” This prompt to Claude then also includes the AAA paper but it drops the reference to the IAA governance framework. To fully get a comprehensive and authoritative list, more research is necessary.
- **Second prompt:** Using these references, list the five key themes of risk, controls, and governance of artificial intelligence in actuarial work, citing specific references and providing links.
- **Result:** Both LLMs contain results consistent with ethical principles articulated in [From Zero to AI Governance](#), with variations. This table includes a comparison.

Claude	ChatGPT	From Zero to AI Governance
Model Governance and Lifecycle Management	Model risk, lifecycle governance, and validation Operational & technical controls (security, vendor/third-party risk, monitoring)	Ensure tools are reliable and safe
Fairness, <b>Bias</b> Mitigation, and Algorithmic Accountability	Fairness, <b>bias</b> mitigation, and <b>explainability</b>	Avoid <b>bias</b>
<b>Transparency</b> and Model <b>Explainability</b>		Make results and uses of AI <b>transparent</b>
Data <b>Privacy</b> and Security	Data governance, lineage, <b>privacy</b> , and quality	Protect <b>privacy</b> and confidential information
Professional Ethics and Actuarial Standards of Practice	Governance, <b>accountability</b> , roles, and proportionality	Ensure there is <b>accountability</b>

- **Observation:** Avoiding bias and protected privacy are key themes across actuarial and industry publications. Explainability or transparency and accountability feature prominently. The LLMs did not specifically reference one principle I articulated to “ensure tools are reliable and safe.” They use model lifecycle management or governance to address this theme.

<sup>4</sup> International Actuarial Association (IAA) Artificial Intelligence Task Force, Governance Workstream, *Artificial Intelligence Governance Framework: General Actuarial Practice: A Comprehensive Governance Framework on Artificial Intelligence Actuarial Work* (Deliverable 2024; draft internal document, PDF), [https://actuaries.org/app/uploads/2025/05/AITF2024\\_G3\\_Governance\\_Framework\\_DRAFT.pdf](https://actuaries.org/app/uploads/2025/05/AITF2024_G3_Governance_Framework_DRAFT.pdf).

<sup>5</sup> Ronald L. Poon Affat, “AI Risk Management Frameworks: An Expert Panel Discussion,” Society of Actuaries Research Institute, April 2025, <https://www.soa.org/resources/research-reports/2025/ai-risk-management-frameworks/>.

<sup>6</sup> Marco De Virgili et al., “Machine Learning in Insurance,” *CAS E-Forum* (Winter 2022), Casualty Actuarial Society, PDF, [https://www.casact.org/sites/default/files/2022-03/01\\_Winter-Eforum-2022-ML\\_in\\_Insurance.pdf](https://www.casact.org/sites/default/files/2022-03/01_Winter-Eforum-2022-ML_in_Insurance.pdf).

<sup>7</sup> National Association of Insurance Commissioners, *NAIC Model Bulletin: Use of Artificial Intelligence Systems by Insurers* (adopted December 4, 2023), PDF, <https://content.naic.org/sites/default/files/cmte-h-big-data-artificial-intelligence-wg-ai-model-bulletin.pdf>.

<sup>8</sup> American Academy of Actuaries, Committee on Professional Responsibility, *Actuarial Professionalism Considerations for Generative AI: A Professionalism Discussion Paper* (September 2024), PDF, <https://www.actuary.org/wp-content/uploads/2024/10/professionalism-paper-generative-ai.pdf>.

**Third prompt:** *Generate a new table, based on the table in the article Zero to AI Governance in the May 2025 edition of the [Actuarial Intelligence Bulletin](#), with redline updates to the current version, and citing specific recommended updates to the principles, standard requirements, or best practices section with references.*

- **Result:** Both LLMs recommend updates to the table, with specific sources cited. Chat GPT gives free form text of one hundred plus lines. Claude generates a table, as requested. The table, labelled *AI Governance Framework: Updated Version with Redline Changes*, is published [here](#).<sup>9</sup>
- **Observation:** The content from Chat GPT is substantive and meaningful, but difficult to present due to the volume. Claude generated a table I was able to publish. This is what I was looking for. My next step is to evaluate the results to update the table from the Zero to AI Governance article.

### Overall Observations

Consistency over time is a challenge. On multiple occasions, using the same prompt across time resulted in different results. This includes the tabulation of resources and the description of the key themes. Prompt engineering matters. To arrive at the three final prompts above, it took iterations to get the table I was looking for from Claude. The more precise and specific the prompt, the better the results. Choice of LLM matters. The results vary by LLM, and had I brought in additional LLMs, they would have varied further. My prompt engineering got the results I wanted from Claude but not from Chat GPT.

AI is a useful and helpful tool in performing research. It has limitations in terms of completeness, consistency, and reliability. Nonetheless, it can save considerable time and produce insights and conclusions that would take much longer for a human to derive. When combined with human oversight and judgment in the form of accountability, one of the key principles articulated in the Zero to AI Governance article, AI can be a valuable research tool.

*Mitch Stephenson, FSA, Head of Model Governance Solutions, Fannie Mae.*



## Can AI Implement Actuarial Model Changes? A Practical Experiment

**IGOR NIKITIN, ASA, MAAA, JOSE SANCHEZ, ASA**

We recently ran a small experiment to test a question many actuaries are quietly asking: can AI make sophisticated actuarial modeling changes, or is it mainly a productivity assistant?

Rather than debating this abstractly, we decided to test it on a real enhancement to an actuarial modeling platform. The task was non-trivial, actuarial in nature, and representative of the kind of work actuary-engineers do every week. We pitted a traditional development approach against an AI-first approach and compared the results.

---

<sup>9</sup> Claude public artifact e85f92c0-06ca-4f9c-8e5e-38761493525e),” Claude, <https://claude.ai/public/artifacts/e85f92c0-06ca-4f9c-8e5e-38761493525e>.

## The Task

The enhancement was to add pre-retirement death benefits to our modeling platform.

These benefits are payable to the surviving spouse of a pension plan participant who dies before retirement. Payments begin when the surviving spouse reaches a specified age and are paid as an annuity.

The actuarial complexity arises from the fact that participants can retire at multiple possible ages and hence the model must track the unretired state accurately. We would classify this as a medium-difficulty actuarial modeling problem.

From a software perspective, the change required edits to five files:

- Four minor updates to register a new benefit type
- One moderately complex update to create a variant of an existing joint annuity class with modified logic for pre-retirement death benefits

## The Competitors

Both approaches were executed by the same person: my colleague Jose Sanchez, an experienced software engineer and actuary.

In the first approach, Jose used his normal development workflow:

- Visual Studio
- GitHub
- GitHub Copilot for in-line code suggestions

The most important lesson was not that AI failed, but how it failed.

In the second approach, his goal was to rely as heavily as possible on AI, using GitHub Copilot Chat and similar tools, to generate the required code.

## The Win Condition

The win condition was strict. The code had to run and pass a defined set of test cases. “Almost right” did not count.

## The Results

Traditional development

Starting at 9AM, Jose reviewed the design documentation, mapped out the required changes, implemented the updates, and ran the tests.

Total time: 6 hours.

This was faster than his initial estimate.

## AI-first Development

The following day, Jose attempted the same enhancement using an AI-first approach.

One advantage was that the requirements were already clear from the prior day. This allowed him to focus specifically on AI's ability to implement the logic he already knew.

The initial attempt involved minimal prompting, asking AI to make the necessary changes autonomously. After roughly an hour, it became clear that this was not working. The AI lacked sufficient understanding of both:

- The actuarial logic behind pre-retirement death benefits
- The structural design of the modeling platform

Next, Jose switched to very detailed prompts that explained the platform architecture and actuarial requirements. This improved results somewhat. The four trivial updates were mostly correct, with minor edits needed. The core benefit class had a reasonable structural skeleton. The actuarial logic of the pre-retirement death benefit was still incorrect

After two more iterations focused solely on the benefit class, the AI produced code that was close but still required minor manual Python edits to fix the logic and pass the tests.

Total time: 8 hours.

## Takeaways

Our debrief led to several clear conclusions:

- For a task of this complexity, AI was unable to produce fully correct actuarial code end to end.
- Writing detailed prompts was time-consuming and cognitively demanding.
- Actuarial calculations generated by AI were frequently “almost right,” which led to multiple iterations of careful review that were time-consuming and mentally straining.
- In-line code completion from GitHub Copilot was consistently helpful and reliable.
- Software structure can be taught to AI effectively, but the actuarial concepts were a struggle even with very detailed prompting.

## Conclusion

For this task, it was faster and far more enjoyable to write the code directly, using AI only as an in-line helper rather than as the primary author. The need to write elaborate prompts, along with multiple iterations of careful review, were the main drawbacks of AI from both performance and work satisfaction perspectives.

The most important lesson was not that AI failed, but **how** it failed. In actuarial work, being “almost right” is often more dangerous than being obviously wrong, because it can survive initial review and slip into production. For now, AI is best treated as a productivity multiplier, not a substitute for an experienced modeler.

*Igor Nikitin, ASA, MAAA, Co-Founder & CEO, Nice Technologies*

*Jose Sanchez, ASA, Co-Founder & Co-CTO, Nice Technologies*



**ACTUARIAL  
INTELLIGENCE  
BULLETIN**

## How One Experiment with Copilot Changed My Entire Workflow

**ARUN PAUL, FSA MAAA**

I used to think AI was something meant for other people—coders, quants, data scientists, anyone who worked in the world of algorithms and models. Definitely not someone like me. My day-to-day work is all about digging through annual statements, reviewing old treaties, and hunting for technical details buried inside long, sometimes barely legible documents. It's meaningful work, but it can be slow and honestly pretty exhausting.

For a while, I assumed AI wouldn't be useful for the kind of messy, unstructured material I deal with. So, I never really bothered with it.

But one day, after spending way too long on a particularly stubborn treaty, I figured—why not try something new?

A small test that changed my mind. I opened Copilot (the AI model available on my company desktop), uploaded the treaty, and typed a simple question:

“What are the key terms of the treaty?”

I wasn't expecting anything impressive. But within seconds, it gave me a clean summary. Not perfect, but surprisingly helpful—especially considering the document I fed it wasn't in great shape. That little result was enough to get me curious. So, I asked more questions.

I asked for the effective date. It found it instantly. I asked about reinsurer risk charges, termination rights and other criteria—things I usually spend half an hour flipping pages for. Copilot pulled them all out and organized them.

Then I wondered whether I could gather the page numbers using AI so I could verify things faster. It worked! That's when something shifted. Suddenly the most tedious part of my job felt different. Instead of combing through hundreds of pages, it felt like a conversation—I asked, it answered.

One treaty was nice. But I usually deal with multiple at once, each formatted differently, each with its own quirks. Under normal circumstances, that means hours of repetitive reading. My goal was to scale this process to streamline the review of as many treaties as possible.

At home, I had been experimenting with AI agents on ChatGPT and saw how easily they could automate repetitive tasks. So, the question popped into my head:

“What if I could build an agent to do this across all my treaties?”

The catch?

At work, the AI tool I have access to is Copilot. And I had never built an agent inside Copilot before. Still, I was in exploration mode, so I asked Copilot how to set one up.

To my surprise, it didn't make me read a technical guide or write code. It walked me through the steps—plain language, simple instructions, nothing intimidating. I followed along, and before long, I had an agent that could read multiple treaties, extract the same key elements, standardize them, and give me a structured view.

What usually took hours now took a fraction of the time and the consistency was better than what I could do manually. That experience didn't just make me faster—it changed the way I think about my work.

I realized AI isn't just for people who build models. It's for anyone dealing with complexity, volume, and information overload—which pretty much describes my job perfectly.

AI didn't replace anything I do—it made me better at it.

It handled the repetitive parts so I could focus more on the actual judgment calls: interpreting terms, understanding implications, evaluating risks. The parts of the job that really matter.

But the biggest shift was letting go of the idea that AI wasn't for me. For years, I thought AI had nothing to offer my type of work. The truth was the exact opposite. AI was exactly what my workflow needed. I just didn't realize it until I finally gave it a chance.

*Arun Paul, FSA MAAA is Vice President, Financial Solutions, SCOR.*



## ACTUARIAL INTELLIGENCE BULLETIN

### Understanding AI Risk Beyond the Model: A Conversation with Karen Bennet

**RONALD POON AFFAT FSA**

**RPA:** Karen, many actuaries will not be familiar with your background or the type of work you do. Before we get into details, could you explain what you work on and why it matters in today's AI-driven world?

**Karen Bennet:** My work sits at the intersection of AI, software engineering, and governance. I help define common standards that allow organizations to clearly describe how AI systems are built, what data they rely on, how they change over time, and what risks they introduce. Earlier in my career I built large-scale software and AI systems directly, including systems used in safety-critical settings. What became clear is that performance alone is no longer the hard problem. The real challenge is trust—being able to explain, audit, and manage AI systems responsibly across technical, business, and regulatory boundaries.

**RPA:** Many actuaries think of AI primarily as “a model trained on data.” What problem are you trying to solve that goes beyond that view?

**Karen Bennet:** The problem is visibility. Modern AI systems are not single models running in isolation. They are collections of components: software, hardware, training data, fine-tuning data, prompts, external services, updates, and operational controls. Yet many organizations cannot answer basic questions about the AI they are using—where it came from, what it depends on, what its limitations are, or how it evolves over time. Without that visibility, AI becomes a black box, and from a risk perspective that is unacceptable.

**RPA:** You often use the example of a self-driving car when explaining this problem. Why is that example so useful?

**Karen Bennet:** A self-driving car is an ideal illustration because no one would seriously claim that its safety depends on a single algorithm. It is a system made up of sensors, hardware, software, machine-learning models, training data, real-time inputs, remote services, and continuous updates. If something goes wrong, you cannot say “the model failed” and stop there. You need to know which sensor misread the environment, which data the system was trained on, how the decision logic worked, and whether an update introduced a new dependency. AI in insurance, healthcare, or finance is becoming similar in structure, even if the physical stakes are different. The lesson is the same: risk lives in the system, not just the model.

**RPA:** Why do these systems view matter so much as AI moves into regulated or high-impact environments?

**Karen Bennet:** Because accountability depends on explanation. In low-stakes settings, opacity may be tolerated. In high-impact settings, it is not. Regulators, courts, and boards increasingly expect organizations to demonstrate how they knew an AI system was appropriate for its use. That means being able to trace decisions back to design choices, data sources, and controls. Without a systems view, organizations are left reacting after failures rather than preventing them.

**RPA:** Let’s make this practical. Suppose an insurer or bank is already using AI tools today, including third-party systems. What should a responsible organization be able to answer?

**Karen Bennet:** They should be able to answer questions like: What is this system intended to do, and what should it not be used for? What data influenced its behavior, and are there known biases or gaps? How often does it change, and who controls those changes? What external services does it rely on, and what happens if those services fail? Has it been tested for safety, fairness, and robustness, and where is that evidence recorded? These are ordinary risk questions. AI simply makes them harder to answer unless documentation is built into the process.

**RPA:** From an actuarial standpoint, we often assess risks arising from systems we did not design. How does better visibility change actuarial risk thinking?

**Karen Bennet:** It allows differentiation. Instead of treating “AI risk” as a single category, actuaries can evaluate governance quality, uncertainty, and control maturity. Two AI systems may use similar techniques but present very

Karen Bennet is a senior AI standards and governance leader with over 40 years of experience in software engineering, open-source infrastructure, and artificial intelligence. A founding member of Red Hat and a former AI startup founder (with two companies acquired by Microsoft), she now focuses on how complex AI systems are documented, governed, and assessed for risk. She plays a leading role in global AI standards development, including chairing and co-chairing multiple initiatives at the Linux Foundation, ISO/IEC, and IEEE, with a particular emphasis on transparency, traceability, bias, and system-of-systems risk in high-impact and regulated environments.

different risk profiles depending on how transparent, stable, and monitored they are. That distinction matters for pricing, capital, vendor selection, and even insurability. Transparency turns AI from an unknowable risk into a structured one.

**RPA:** Actuaries think naturally in terms of lifecycle risk. How should we think about the lifecycle of AI systems?

**Karen Bennet:** AI systems do not stop evolving at deployment. Data changes, models drift, dependencies update, and usage expands beyond original intent. Some systems even learn continuously. Risk therefore changes over time. Organizations need to track those changes and detect when assumptions no longer hold. Without lifecycle awareness, decisions are based on outdated information, which is dangerous in any risk-sensitive context.

**RPA:** You've also warned about increasing autonomy in AI systems. Why does that raise governance concerns?

**Karen Bennet:** As systems become more autonomous, human oversight becomes less immediate. Decisions happen faster and with fewer manual checks. That increases the importance of upfront design discipline, clear boundaries, and ongoing monitoring. You cannot rely on humans to catch every issue after the fact. Governance must be built into the system itself.

**RPA:** Some readers may worry that this level of documentation is unrealistic. Is this achievable?

**Karen Bennet:** It must be automated, or it will fail. Manual documentation does not scale. The goal is structured information that is generated as systems are built and updated, not paperwork added at the end. When done properly, documentation becomes a tool for managing risk rather than a compliance burden.

**RPA:** Finally, for actuaries or risk professionals who want to learn more or get involved, where should they start?

**Karen Bennet:** I would encourage them to engage with open standards communities and AI governance working groups. These forums publish their work openly and actively welcome input from risk and insurance professionals. Actuaries bring a critical perspective because they understand uncertainty, tail risk, and long-term consequences. Even asking the right questions can shape how these frameworks evolve.

*Ronald Poon-Affat FSA, FIA C.Act, MAAA, CFA, HIBA is an Independent Board Member, and Senior Consultant rpoonaffat@soa.org*



## ACTUARIAL INTELLIGENCE BULLETIN

### Why AI Can Systematically Amplify Risk — and What Actuaries Can Do About It

**MARC SCHMITT, PHD**

Artificial intelligence is now deeply embedded in how modern risk is measured and managed. From market models and capital calculations to operational risk and cyber defense, AI systems increasingly inform decisions that were previously made by humans or governed by static rules. Much of the discussion has focused on whether these systems are accurate, explainable, and subject to responsible governance. This focus leaves a related issue largely unaddressed.

Trading and forecasting algorithms learn from one another's behavior. When many systems react to similar patterns—such as momentum signals, volatility changes, or liquidity conditions—their individually rational actions can become synchronized. During periods of stress, this synchronization can magnify price movements and contribute to liquidity withdrawal.

AI does not only improve the efficiency of risk measurement. Under certain conditions, it can systematically amplify risk, even when individual models are well designed and behave rationally.

Risk is no longer only exogenous. It can emerge endogenously from the interaction of intelligent systems themselves.

### From Better Models to Risk Amplification

Traditional actuarial thinking treats volatility, losses, and extreme events as responses to external shocks such as economic downturns, catastrophes, behavioral shifts, or rare tail events. Models are designed to absorb these shocks, estimate their likelihood, and maintain resilience through capital buffers and diversification.

AI changes this setting. In modern financial and operational systems, models often do not operate in isolation. They are trained on overlapping data, react to similar signals, and adapt to one another's outputs. Trading algorithms respond to prices set by other algorithms. Risk systems incorporate signals generated by AI-driven information platforms. Forecasting models are retrained on outcomes that earlier models helped produce.

When learning systems observe and adapt to one another, feedback loops form. These loops can turn small disturbances into persistent volatility, clustered losses, or sudden breakdowns, even in the absence of new fundamental information.

This can be described as algorithmic reflexivity: AI systems increasingly shape the environment from which they learn.

### Algorithmic Exuberance: A Structural Risk Mechanism

In my recent work on [Algorithmic Exuberance](#), I formalize this mechanism in the context of financial markets. Volatility can arise endogenously from the interaction of two feedback channels operating within AI-driven systems.

The first is market-side reflexivity. Trading and forecasting algorithms learn from one another's behavior. When many systems react to similar patterns—such as momentum signals, volatility changes, or liquidity conditions—their individually rational actions can become synchronized. During periods of stress, this synchronization can magnify price movements and contribute to liquidity withdrawal.

The second is information-side reflexivity. Information systems, including news aggregation, sentiment analysis, recommendation engines, and generative AI, amplify and redistribute market signals. Price movements generate attention, narratives, and sentiment, which are then incorporated into trading models as new inputs.

Together, these channels form a closed feedback loop. Prices influence information, information influences models, and models influence prices. The system becomes self-referential.

This structure is consistent with more persistent volatility and fatter tails in modern markets relative to earlier periods, despite faster and more sophisticated information processing. Information may be incorporated quickly, while instability persists in the dynamics.

The central implication is not specific to markets. It concerns how interacting intelligent systems can reshape the risk environment.

### Human-Centered Risk: Digital Deception

The amplification mechanism described above is not confined to asset prices. Similar feedback dynamics appear in other risk domains.

A related amplification mechanism appears in human-centered risk. In my joint work on [Digital Deception](#), we examine how generative AI transforms social-engineering and phishing attacks. The change is not limited to more realistic content. It alters the structure of these attacks.

Generative AI enables realistic content creation, advanced targeting and personalization, and automated attack infrastructure. Human beliefs, trust, and behavior become part of a feedback process. Successful attacks generate data that improve subsequent targeting. Narratives can be adapted quickly in response to defensive measures.

Risk amplification arises through this recursion. Loss frequency and severity are shaped not only by external adversaries but by interaction between attackers, defenders, and AI systems that mediate communication.

Across markets and human behavior, the pattern is similar. AI can shift risk from something largely external to something generated within the system itself.

### Why This Matters for Actuaries

This shift has practical implications for actuarial work.

First, diversification assumptions can weaken when models learn from similar data and architectures. Correlations can rise during stress, when diversification is most relied upon.

Second, stress testing and scenario design should account for adaptive responses. Traditional scenarios assume fixed reactions. In AI-driven settings, reactions can change during the stress event itself, which can amplify losses.

Third, model risk expands from individual models to interacting systems. The risk is not only that a model is incorrect, but that multiple models that perform well in isolation interact in destabilizing ways.

Finally, risk categories can blur. Market, operational, cyber, and behavioral risks may share amplification mechanisms. Treating them in isolation can understate systemic exposure.

### What Actuaries Can Do

The aim is not to avoid AI. In many settings, intelligent systems are already integral to risk work. The relevant challenges concern governance, design, and measurement.

Three practical considerations follow.

1. Model interaction matters alongside model accuracy.  
Validation frameworks should consider how models behave under stress when multiple systems respond to similar signals, not only how a model performs in isolation.
2. Feedback intensity can be monitored.  
Indicators of persistence, clustering, and regime shifts can provide signals of feedback-driven dynamics and help identify transitions toward amplification-dominated regimes.

3. Human judgment remains a stabilizing component. Human oversight introduces heterogeneity, friction, and context. These features can dampen synchronization and reduce the likelihood of runaway amplification.

The objective is not to eliminate volatility, but to limit self-reinforcing amplification.

## A Closing Thought

AI has made our systems faster, smarter, and more adaptive. But intelligence has a side effect: it generates unintended results.

When learning systems interact, risk can be generated through feedback within the system.

Actuarial work has long focused on systemic interactions, tail behavior, and resilience over time. These perspectives remain applicable in settings where risk increasingly reflects feedback and interaction rather than isolated shocks.

The practical challenge is not to make AI risk-free. It is to ensure that adaptive systems remain stable under feedback and do not amplify their own signals.

This defines the task of risk management in AI-enabled environments.

## References

Schmitt, M. (2025). *Algorithmic Exuberance*. SSRN Working Paper. <https://doi.org/10.2139/ssrn.5742643>

Schmitt, M., & Flechais, I. (2024). *Digital deception: Generative artificial intelligence in social engineering and phishing*. *Artificial Intelligence Review*, 57(12), 324. <https://doi.org/10.1007/s10462-024-10973-2>

Marc Schmitt PhD, University of Oxford



**ACTUARIAL  
INTELLIGENCE  
BULLETIN**

## Can I Trust My AI Output? Prompt Design, Variance, and Validation of LLM-driven Insurance Systems

**JEFF HEATON, PH.D.**

Large language models (LLMs) are increasingly becoming standard components of modern enterprise software systems at insurance companies. They can be found in agentic systems that operate within governed enterprise environments, supporting decision processes while ensuring auditability, explainability, and risk controls. Agentic software can understand a situation, decide what to do next, and take action to achieve a goal, often with limited human input. These agentic systems have been adopted across underwriting, claims management, fraud detection, customer service, policy administration, and compliance monitoring.

LLMs, such as Claude, Llama, GPT, and others, typically give slightly different answers to questions each time, depending on their configuration. This is like asking the same question to a human several times over the course of a day, when answers may vary. In addition, the companies behind these LLMs usually update their models periodically. These factors can make it difficult to evaluate the accuracy and consistency of LLM-dependent

computer systems over time, as output-only comparisons are often insufficient to ensure that system behavior has not meaningfully changed.

This article looks at how insurance companies audit and test systems built on constantly changing LLMs that, even without a version upgrade, already produce inconsistent results.

### Dealing with Inconsistent Results

Enterprise systems that use LLMs must be designed to mitigate the effects of inconsistent results. Companies should test these systems at two levels: the LLM’s actual outputs and the enterprise system’s overall accuracy.

Starting with overall accuracy, prompts should be designed to mitigate variance in answers as much as possible. Consider the following poorly conceived prompt:

*Produce a one-sentence summary of life insurance underwriting concerns for the following medical document.*

Running this prompt on a synthetic medical dataset produced the following two outputs in response to two subsequent queries:

- *A 50-year-old male presents with moderately increased underwriting risk due to uncontrolled hypertension, early type 2 diabetes (A1c 6.6%), overweight status, and suboptimal medication adherence, but with preserved renal function, no cardiovascular events, and a favorable non-smoking history.*
- *A 50-year-old male presents with life insurance underwriting concerns due to uncontrolled hypertension with inconsistent medication adherence, newly diagnosed early type 2 diabetes (A1c 6.6%), overweight status, dyslipidemia, sedentary lifestyle, and a family history of cardiometabolic disease, collectively increasing long-term cardiovascular risk despite currently normal renal function and no end-organ damage.*

Making the prompt more precise produces much more consistent results. The following prompt requests a list of medical codes and the top concerns associated with them:

*Extract the most important life insurance underwriting concerns from the medical document. Output only CSV with three fields per row: Concern, ICD-10 Chapter, Severity (1–10). Use standardized medical terms, rank by severity (highest first), and include only documented conditions or risk factors.*

This prompt is considerably more robust. It describes exactly what is sought and the precise form in which it should be delivered. The precise prompt has several benefits, including more consistent results in the form of a simple list with three columns: a medical code, the name, and the LLM’s assessment of severity.

Severity scores likely will contain the most variance. In practice, insurers likely will have an entire prompt dedicated to this severity ranking with very detailed expressions – if they even trust LLMs to rank severity at all. That said, if using only the data and not the assessment, severity scores still can help rank the LLM-extracted data.

From here, testing the system for accuracy can proceed more efficiently, focusing less on the output of individual LLM queries.

A relatively new approach to measuring response similarity is to simply ask the LLM how similar two responses are. While this may seem like “asking the fox to watch the hen house,” it has become one of the most common approaches.

## Do Two Different Answers Mean the Same Thing?

Determining the similarity of two responses requires testing an LLM's actual output. From a quick read-through, the two responses in the previous section – elicited by an ill-conceived prompt – appear quite similar. However, some important, although relatively minor, differences become evident upon closer inspection.

The key is to apply a method to compare LLM responses and quantify their differences. Rather than simply comparing individual words, several mathematical approaches can be used. These include measuring the degree of semantic overlap between the two responses using vector similarity techniques or checking whether one response logically implies or contradicts the other.

A relatively new approach to measuring response similarity is to simply ask the LLM how similar two responses are. While this may seem like “asking the fox to watch the hen house,” it has become one of the most common approaches. Consider this prompt:

*Compare the following two LLM responses. Focus only on semantic meaning and factual content. Ignore wording and style. Note any clinically relevant information that is added, missing, or changed.*

*Output exactly:*

*[similarity score 0–10, where 10 = identical], one sentence describing the key difference(s). Do not add commentary.*

For the same two responses examined above, this prompt produced the following:

*7, Response B adds dyslipidemia, sedentary lifestyle, and family history while omitting the favorable non-smoking history and explicit absence of cardiovascular events noted in Response A.*

How could two responses, both for the same model version, produce a score of only 7/10? While this might be quickly chalked up to hallucination, something else is going on: The original imprecise prompt is the culprit.

The second, improved prompt, consistently produced results with a score of 9 or 10. If prompts are imprecise, resulting in a 7/10 variance, it will be very difficult to test in the future and to evaluate the variance introduced by an entirely new model/version.

These types of variance comparisons are important when designing prompts, and output should be regularly evaluated to determine potential need for fine-tuning.

## Conclusion

LLMs can be powerful tools within a controlled insurance system, but their inherent variability means accuracy, consistency, and auditability must be engineered into the system rather than assumed from the model. Precise prompt design and systematic variance testing are essential foundations for reliably evaluating LLM-based systems over time, especially as models evolve and are replaced.

*Jeff Heaton, Ph.D., VP, AI Innovation, RGA*

## SOA Expert Panel Report: The AI-Ready Actuary

**RONALD POON AFFAT, FSA, FIA C.ACT, MAAA, CFA, HIBA**

**AIB:** For this report, you gathered perspectives from academia, insurance, reinsurance, brokerage, and consulting. Why do you think that the panel decided that focusing on mindset – not technical skills like coding – was the most urgent priority for the profession right now?

**Ronald Poon Affat:** The decision was very intentional. While technical proficiency is necessary, it is a moving target. The specific AI tools and languages we discuss today will evolve or be replaced. What remains constant is the need for sound professional character. Our central inquiry was: As AI becomes embedded in our workflows, what are the timeless human qualities that will ensure actuaries continue to add essential value? We needed to move beyond the noise about the latest model and focus on the durable blend of skepticism, ethical judgment, and clear communication that defines our profession’s credibility, especially when computational speed is no longer a differentiator.

An AI can optimize for statistical precision or profit metrics. An actuary must contextualize those outputs within a broader human and societal framework. We evaluate fairness, assess broader impact, and consider the real-life implications for policyholders and claimants. This is the profession’s bedrock value.

**AIB:** The report emphasizes concepts like "disciplined thinking" and "problem framing." Could you illustrate what **this looks like in a concrete, day-to-day scenario for an actuary reviewing an AI’s work?**

**Ronald Poon Affat:** Certainly. Imagine an AI system proposes a radically new underwriting segmentation that promises significant profit improvement. The traditional actuary might jump to validating the calculations. The AI-ready actuary, however, begins by reframing. They step back to ask: “What specific business challenge were we addressing with this analysis? What are the defined criteria for a successful and responsible outcome?” This is the shift from *calculator* to *orchestrator*. They then apply "informed skepticism," pressure-testing the AI’s compelling output against core principles: Does this align with our historical loss experience across all segments? Does it logically hold for different demographic groups? Their critical contribution becomes asking the probing, clarifying questions that the AI, by its nature, is not designed to ask.

**AIB:** A foundational principle in the report is that AI should be treated as "informative rather than authoritative." This seems particularly crucial when explaining results to non-technical executives. How does an actuary put this principle into practice in those high-stakes conversations?

**Ronald Poon Affat:** This is where an actuary’s role as a communicator and guide becomes paramount. It requires the ability to build a narrative that demystifies the AI’s role. An actuary must be prepared to explain: “The AI identified these patterns in the data, which is informative. Our professional responsibility was to then interpret those patterns. Here are the business assumptions we reviewed, the ethical guardrails we considered, and the practical safeguards we implemented based on our judgment.” The principle reinforces that we never outsource accountability. As a panelist from the consulting space emphasized, our authority has always stemmed from our judgment. AI makes the transparent communication of that judgment—its rationale and its limits—a non-negotiable core competency.

**AIB:** One finding states that "peer review retains central importance." Given that peer review is already a professional standard, how does its function change or intensify in an AI-augmented environment?

**Ronald Poon Affat:** Its function evolves from verification to rigorous investigation. When analyses are generated almost instantaneously, the focus of peer review must shift upstream from the *answer* to the *process that created it*.

It becomes a collaborative audit. Peers now need to interrogate: “What was the provenance and quality of the training data? What optimization goal was programmed into the model? How did we test for unintended bias or ‘hallucination’ in the results?” This transformed peer review acts as the essential human governance layer—a structured, collective exercise in applying professional scrutiny to automated processes. It is our most robust defense against uncritical adoption of AI-generated outputs.

**AIB:** The panel underscored that "AI cannot interpret the human dimensions" inherent in actuarial work. What specific human qualities become the actuaries' irreplaceable toolset for providing this interpretation?

**Ronald Poon Affat:** The essential qualities are empathy, contextual intelligence, and ethical foresight. An AI can optimize for statistical precision or profit metrics. An actuary must contextualize those outputs within a broader human and societal framework. We evaluate fairness, assess broader impact, and consider the real-life implications for policyholders and claimants. This is the profession’s bedrock value. The AI-ready actuary brings a necessary humility and human perspective to the table, consistently asking questions beyond the algorithm’s scope: “Is this outcome equitable? Could this create unintended barriers for certain groups? Does this decision align with our fiduciary and social responsibilities?” This layer of human judgment is what transforms data-driven output into responsible business practice.

**AIB:** The report also highlights the need to modernize "apprenticeship" for early-career actuaries. What is the core adaptation required in how we train new professionals?

**Ronald Poon Affat:** The core adaptation is a deliberate shift from training focused on manual calculation to training centered on building critical judgment. If AI performs the computational heavy lifting, we must create explicit, structured opportunities for early-career actuaries to develop intuition and reasoning. Mentors should facilitate “judgment gyms”—practical sessions where trainees deconstruct AI-generated reports, challenge the underlying assumptions, and practice articulating the *why* behind conclusions. The objective is to cultivate a generation of actuaries who are not just users of technology but its informed, confident, and critical stewards, thereby preventing an over-reliance on a "black box."

**AIB:** Finally, synthesizing the panel’s insights, what is the single most important message for any actuary looking to navigate this transition successfully?

**Ronald Poon Affat:** The paramount message is to cultivate the mindset of the responsible orchestrator. Your enduring professional value lies not in computational speed but in your seasoned judgment, your ethical reasoning, and your unparalleled ability to interpret complexity and communicate actionable insights. Invest deliberately in deepening the intrinsically human skills that form our profession’s cornerstone: principled skepticism, nuanced communication, and an unwavering commitment to those impacted by our work. Ultimately, hone your ability to ask the more insightful question. That is the capability that remains forever beyond automation’s reach.

*Ronald Poon-Affat FSA, FIA C.Act, MAAA, CFA, HIBA is an Independent Board Member, and Senior Consultant rpoonaffat@soa.org*



**ACTUARIAL  
INTELLIGENCE  
BULLETIN**

## AI Will Not Replace Actuaries, but Actuaries Who Adapt Will Replace Those Who Don't

**SYED AZADAR HAIDER, FSA, FPSA**

I opened a blank document to begin writing this article and found myself staring at the blinking cursor longer than I expected. Not because I lacked an opinion on artificial intelligence, but because a different thought briefly crossed my mind: *I could probably have this drafted in a few minutes if I let AI do it.*

It was a fleeting moment, almost unnoticeable. Still, it felt familiar. Many of us experience the same pause when starting a model, reviewing assumptions, or preparing an explanation for management. Efficiency quietly pulls us in one direction, while professional ownership pulls us in another. I chose to write this piece myself, not as a rejection of AI, but as a reminder that originality and judgment still matter.

Around five years ago, predictive analytics was formally introduced into actuarial education, with concepts such as neural networks discussed largely at a theoretical level. At the time, it was presented as an emerging area, something students should be aware of, not necessarily something they would use immediately. There were references to experimental work where organizations were exploring claim assessment using images. I recall not giving this much thought.

Fast forward to today, and within a relatively short span, we have seen tangible progress across insurance, medical sciences, and several other fields. Claims are processed with minimal human intervention, algorithms assist in surgeries, and systems suggest first-line treatments. Few would have predicted that such developments would move from theory to practice so quickly.

Looking further back, when I started my career over a decade ago, skills such as VBA in Excel or basic Python coding were considered a significant advantage. Over time, that “advantage” gradually became a minimum expectation. Today, many candidates, despite not having a strong formal background in programming, are able to code at a very early stage of their careers, largely due to the availability of modern tools and learning resources. This observation is not meant to undermine the value of knowledge or effort; rather, it highlights a reality that access to better resources allows individuals with relatively basic foundations to progress faster than ever before.

Being a good actuary does not mean performing every calculation personally. It means ensuring that everything produced, whether by a model, a script, or an AI system, passes through disciplined thinking and professional scrutiny. Machines can process data, identify patterns, and run scenarios faster and often more consistently than we can. That is precisely why they should be used. The value of the actuary lies in interpretation, challenge, context, and communication, areas where judgment, experience, and accountability cannot be automated.

History has consistently shown that the pace of change only accelerates. While it is unrealistic to master every new development, there is little downside in moving forward with technology rather than against it. In a fast-changing world, actuarial thinking will not become obsolete. But actuaries who refuse to adapt risk doing so themselves.

*Syed Azadar Haider, FSA, FPSA, Senior Manager, BADRI*

Being a good actuary does not mean performing every calculation personally. It means ensuring that everything produced, whether by a model, a script, or an AI system, passes through disciplined thinking and professional scrutiny.

**ACTUARIAL  
INTELLIGENCE  
BULLETIN**

## The Pilot and the Plane: Actuary-in-the-Model

**DAVE INGRAM, FSA, CERA**

The dominant metaphor for large language models today is the “copilot.” It sounds sensible and reassuring. A copilot is capable, alert, and helpful, yet subordinate to the Actuary in charge. The metaphor implies partnership without surrender and assistance without abdication. It suggests that as long as an Actuary remains “in the loop,” control is preserved.

Unfortunately, metaphors matter, and this one quietly misleads. In practice, the copilot framing shifts cognitive agency away from the Actuary and toward the machine in ways that are subtle, cumulative, and increasingly difficult to detect. The model suggests. The Actuary reacts. Over time, judgment migrates downstream.

In high-stakes domains—medicine, law, public policy, finance, actuarial science, engineering, national security—this shift is not benign. Errors are not inconveniences; they are harms. In these settings, the central risk is not merely that language models can hallucinate or make mistakes. The deeper risk is that they do so while presenting outputs that feel complete, confident, and internally coherent, encouraging passive acceptance rather than active judgment.

The Actuary-in-the-Model proposal begins by rejecting the copilot metaphor entirely.

The LLM is not the copilot here.

It is the plane.

This is not wordplay for rhetorical effect. It is a structural redefinition of roles. A plane is powerful, complex, and indispensable—but it does not decide where to go. It does not choose altitude, route, or destination. It does not reinterpret instructions mid-flight. It executes, precisely and obediently, under the continuous control of a pilot who remains intellectually engaged.

In Actuary-in-the-Model, the LLM is the plane: an extraordinarily capable instrument for executing reasoning steps. The Actuary is the sole pilot. There is no shared agency and no implied autonomy. The Actuary does not “work with” the model’s reasoning; the Actuary commands it.

This distinction matters because most current safety approaches still assume the model will think first and the Actuary will review later. That is oversight. It is not control. Reviewing a finished answer is the cognitive equivalent of reading the flight log after landing. By then, the important decisions have already been made.

Actuary-in-the-Model prevents that inversion by embedding the Actuary directly inside the reasoning process itself. The model does not advance unless instructed. It does not reason freely and await correction. It executes only the step it is told to execute, then stops.

The protocol begins with something deceptively simple: the explicit selection of a formal thinking mode. The Actuary chooses a named, well-established cognitive framework—actuarial reasoning, systems thinking, critical thinking, probabilistic reasoning, first-principles analysis. These are not ad-hoc prompt styles; they are intellectual traditions with known disciplines, strengths, and blind spots.

This step matters because it makes an invisible choice visible. Every serious problem can be approached in multiple legitimate ways, particularly problems about the future. Choosing a thinking mode is an act of judgment. When left implicit, that judgment is silently delegated to the model's training mix. When made explicit, it becomes inspectable, challengeable, and repeatable.

Selecting the thinking mode is equivalent to choosing the flight destination.

The second step is defining the exact sequence of reasoning steps the model must follow. These steps operationalize the chosen thinking mode. Instead of asking the model to "analyze the problem," the Actuary specifies what analysis means in this context and in what order it must occur.

This is the flight plan.

The model is then instructed to execute only the first step. It produces the result and halts. No extrapolation. No anticipation. No helpful next steps. It waits.

At this point, the protocol imposes its most distinctive requirement: mandatory Actuary probing at every gate. The Actuary must ask at least one substantive question—clarifying, challenging, or deepening—about the step just completed. Passive approval is not permitted. Advancement requires interrogation.

This requirement is often misunderstood as a safety check. It is better understood as a discipline of engagement. Asking questions forces the Actuary to remain cognitively present. It prevents the quiet slide from pilot to passenger.

Only after the Actuary is satisfied—or revises the step—does the model receive authorization to proceed to the next step. The cycle repeats until the reasoning process is complete.

The result is a full, auditable transcript of reasoning, questions, revisions, and decisions. Every assumption is surfaced. Every fork in the road is visible. Nothing happens "inside the model" that the Actuary has not explicitly directed or approved.

Seen through the plane analogy, the protocol's strictness becomes intuitive rather than heavy-handed. A pilot does not relinquish the controls because the aircraft is sophisticated. Sophistication increases the need for disciplined control; it does not reduce it.

This approach also clarifies what Actuary-in-the-Model is not. It is not chain-of-thought prompting, where the model is encouraged to reveal its internal reasoning. It is not process supervision layered on top of autonomous output. It is not collaboration in the sense of shared agency.

Actuary-in-the-Model is command.

One of its practical advantages is that it requires no new models, retraining, or proprietary infrastructure. It works with existing LLMs today. This shifts the governance problem from model design to model use. Institutions do not need to wait for safer copilots; they can impose safer flight rules immediately.

There is also a less obvious benefit. Over time, this protocol trains the Actuary as much as it constrains the model. Users become better at selecting appropriate thinking modes, better at formulating probing questions, and better at recognizing where judgment is being exercised. The system accelerates judgment development rather than eroding it.

That matters because the greatest long-term risk of advanced language models is not that they will be wrong. It is that Actuaries will become less practiced at noticing when judgment is required.

The future of trustworthy AI does not require smarter copilots.

It requires planes that do exactly what they are told—and Actuaries who remain firmly in the cockpit, directing every move, challenging every step, and never mistaking power for agency.

*Dave Ingram, FSA, CERA, Managing Editor, Actuarial Intelligence Bulletin*



**ACTUARIAL  
INTELLIGENCE  
BULLETIN**

## Editorial Committee

*Jon Forster, ASA*

*Dave Ingram, FSA*

*Jing Kai Ong, FSA*

*Ronald Poon Affat, FSA*

*Frank Quan, PhD*

*Lisa Schilling, FSA*

*Darryl Wagner, FSA*



Thank you for reading the March 2026 SOA Research Institute AI Bulletin. We hope you found these insights valuable. Stay tuned for future editions as we continue to explore the evolving landscape of AI and its impact on the actuarial profession. We encourage you to engage with the SOA Research Institute and share your own experiences and perspectives on AI. For questions, comments, and article submissions, contact [rpoonaffat@soa.org](mailto:rpoonaffat@soa.org).

## About the Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors, and non-governmental organizations, building an effective network which provides support, knowledge, and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops, and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its [strategic research programs](#): aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of [topical research available](#), including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute  
8770 W Bryn Mawr, Suite 1000  
Chicago, IL 60631  
[www.SOA.org](http://www.SOA.org)