

ACTUARIAL INTELLIGENCE BULLETIN



May 2026

AI in the Actuarial C Suite: Cost and Capability	2
The Actuarial Efficacy Paradox: Why AI Feels Fast but Works Slow.....	4
Case Study: GenAI-Driven Market Comparison	6
AI as a Second Opinion in Claims Reserving	9
Map, Measure, Govern, and Manage: An Actuarial View of the Financial Services AI Risk Management Framework.....	12
Stages of AI Adoption for Businesses	14
From Drafts to Done: Using AI to Streamline Communication	16
Building Fair and Accurate Predictive Models: An Interdisciplinary Imperative for Modern Insurance	19
NIST Launches Trustworthy AI Initiative for Critical Infrastructure	21
Insurance for AI Makers	23
Every Voice Matters: How AI-Powered Voice Translation is Improving Insurance Claims Operations	27
The Rise of the High Throughput Operator	29
Editorial Committee	31
About the Society of Actuaries Research Institute	32

Welcome to the May 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 in the Actuarial C Suite: Cost and Capability

DAVE DILLON, FSA, MAAA, FCA

What Senior Leaders Should Ask as AI Becomes Embedded in Actuarial Work

Artificial intelligence (AI) has moved quickly from "interesting demo" to "Board level question." For actuarial organizations, the most important conversation is no longer whether AI can draft text, write code, or summarize data. The real question is how AI changes expectations, controls, talent development, and accountability for conclusions that influence pricing, reserving, capital management, and strategic decisions.

Two practical starting points for leaders are: (1) the full cost of implementation (including the hidden time cost of training and workflow redesign) and (2) how to develop younger actuaries when AI can perform many early career tasks. Those are essential, but they are not the full story. At the executive level, AI also introduces new governance demands, model risk considerations, cybersecurity exposure, and client and regulator expectations.

This article offers a leadership lens on AI's structural implications: the true cost of implementation, the governance required to manage new risks, and how executive teams can embed AI into actuarial workflows without compromising professional standards.

Points to Ponder

If AI cuts a task from half a day to an hour, will stakeholders keep expectations constant - or will service level commitments quietly tighten?

If a regulator or auditor asks, "Why should I trust this result?" what evidence can the team provide beyond the AI output itself?

1. The True Cost of Implementation: the Visible Budget and the Hidden Calendar

Most AI business cases begin with licensing costs and potential efficiency gains. Those are measurable, but they are often the smallest part of the investment required to make AI safe and useful in actuarial workflows.

Hard dollar costs commonly include:

- Enterprise tools (licenses, usage-based fees, hosting, and model access).
- Implementation support (integration, identity and access management, model operations tooling).
- Data engineering (cleaning, lineage, labeling, storage, and availability).
- Security and compliance (data loss prevention controls, monitoring, audit logs, vendor due diligence).
- Legal and procurement time (contracts, IP terms, indemnities, retention requirements).

Soft costs are frequently larger and more persistent. They show up as calendar time, leadership attention, and temporary reductions in productivity while teams learn to use AI responsibly.

Common soft costs include:

- Training time and early rework as teams learn what AI can and cannot do reliably.
- Workflow redesign (you do not "add AI"; you re-architect steps and review points).
- Higher Quality Assurance (QA) and review burden in the early stages to maintain quality and defensibility.
- Change management (policies, norms, and trust-building across stakeholders).

2. Governance and Accountability: Who Owns the Outcome When AI Contributes?

Actuaries are accustomed to standing behind their work. AI does not reduce accountability; it concentrates it. Executives should set a clear principle: AI is a tool, not an author of record. Human professionals remain responsible for appropriate use, review, and communication of any AI assisted output.

A workable governance approach usually includes:

- Approved use cases and prohibited uses (especially around confidential or regulated data).
- Human in the loop requirements for high stakes outputs (pricing indications, reserve analyses, capital results).
- Documentation standards and review protocols (what must be checked, by whom, and how it is recorded).
- Escalation paths and stop rules (what triggers additional review, rollback, or tool restriction).
- Disclosure norms (internal and external) for when AI was used and how outputs were validated.

The governance goal is not to eliminate risk; it is to make AI use predictable, controllable, and auditable, consistent with the organization's risk appetite and professional obligations.

3. Model Risk Management: Auditability, Explainability, and "Right for the Right Reasons"

AI changes the risk profile of actuarial work because it can produce outputs that sound confident while being wrong, incomplete, or poorly grounded. In traditional actuarial modeling, teams can often trace an answer back to inputs, assumptions, code, and review notes. With AI, traceability must be designed into the workflow.

Common failure modes that matter to leaders include:

- Hallucination and fabricated citations in narrative content.
- Silent errors in code generation or spreadsheet logic that appear plausible.
- Overconfidence bias: output tone masks weak assumptions or missing limitations.
- Inconsistent results when prompts, model versions, or context change.

A practical executive stance is to require stronger documentation for AI assisted work, not weaker. "The model said so" is not defensible to a regulator, auditor, Board, or client. AI assisted analysis should still be supported by clear assumptions, reasonableness checks, and transparent limitations.

4. Data, Privacy, and Cybersecurity: the Fastest Route to Leakage is Convenience

Actuarial teams handle sensitive data, including PHI, PII, claim details, provider arrangements, underwriting information, and strategic assumptions. Generative AI expands the risk of unintended disclosure because it is easy to paste content into a tool to "speed things up."

Executive considerations include:

- Data classification and access controls (what can be used where, and by whom).
- Approved tool environments that keep data inside controlled enterprise boundaries.
- Vendor risk management (retention, training terms, location of processing, subcontractors).
- Monitoring and incident response readiness for AI related leakage scenarios.

A mature posture does not rely on "please be careful." It provides an approved path that is easier than the unsafe alternative.

5. Practical Next Steps: an Executive Starter Playbook

AI becomes a durable advantage when leadership treats it as a capability with clear outcomes, guardrails, and measurement. A pragmatic starting playbook looks like this:

- Select 3 to 5 high value use cases tied to cycle time, quality, growth, or risk reduction.
- Stand up governance: approved tools, permitted uses, review standards, and escalation paths.
- Invest in safe infrastructure: identity controls, data loss prevention, logging, and vendor due diligence.

These investments build the structural foundation. But infrastructure alone does not determine whether AI strengthens or weakens actuarial credibility. That question depends on how results are communicated, what behaviors the organization rewards and how people are trained.

Dave Dillon, FSA, MAAA, FCA is Chair & President – Society of Actuaries and SVP & Principal – Lewis & Ellis, LLC.



ACTUARIAL INTELLIGENCE BULLETIN

The Actuarial Efficacy Paradox: Why AI Feels Fast but Works Slow

SHANE LEIB, FSA, MAAA

The practical promise of AI has long seemed simple. Even if the boldest claims about superintelligence or mass automation remain distant, AI should at least make knowledge workers more productive. The 2025 METR study disrupted that assumption. It studied experienced open-source developers working on real issues in large repositories they knew well. Participants expected AI to speed them up by 24%. Even after the work was done, they still believed it had helped improve by 20%. Yet the measured result showed they were slower when AI tools were allowed by 19%.

That finding matters beyond software engineering. It matters for actuarial modeling because the study exposes a larger truth: AI often helps most where work is shallow, structured, and easy to verify, but it can become a drag where work is expert, contextual, and accountability-heavy, exactly the kind of work actuaries do.

The force of the METR result lies in the setting. This was not a benchmark or a classroom exercise. These were real tasks in mature codebases, performed by people with deep familiarity and high standards. The measured outcome was not output volume or perceived helpfulness, but end-to-end completion time. That distinction matters because expert work is rarely constrained by typing speed. More often, it is constrained by judgment, integration, validation, and defensibility.

Actuarial modeling shares those same constraints, creating a unique efficacy paradox. While AI provides instant momentum by generating drafts or code, the real bottleneck isn't production, it's professional judgment. Because an actuarial task is only "complete" once it is validated, reconciled, and made defensible to regulators, speed at the front end often simply reappears as extra effort at the back end.

Why AI Slows Down Expert Work

Two forms of debt are especially important here. The first is verification debt: the additional effort required to prove that AI-assisted output is correct. The second is defensibility debt: the additional effort required to explain why that

output is correct in a setting where explanation matters. In actuarial work, both debts are expensive because the standard is not merely to produce something plausible, but to produce something reliable and accountable.

If anything, actuarial modeling is even more vulnerable to these problems than software development. Actuarial systems are full of implicit context. Product features, rider interactions, tax rules, accounting treatments, policyholder behavior, and long-standing control logic are often only partly documented. Much of the real knowledge sits in institutional memory. The METR study found that AI struggled when important knowledge was implicit rather than explicit. That problem is likely more severe, not less, in actuarial environments.

Additionally, there is silent error risk. In many domains, an error announces itself by breaking something. In actuarial modeling, an error may still produce a result that looks reasonable. A reserve can be wrong without looking absurd. A projection can seem fine until it fails in a downstream reconciliation. That makes “almost right” especially dangerous. In actuarial work, a plausible answer is not enough.

A third reason is timing. Problems in actuarial modeling often surface late. They show up in validation, management reporting, audit, or regulatory review, long after the original output was produced. By then, a small error may have spread across exhibits, explanations, and decisions. AI therefore does not simply create a technical risk. It can create a lagged governance risk.

And over all of this sits professional accountability. The actuary owns the work product. If the result is signed, explained, or relied upon, responsibility remains with the human professional, whether the underlying work came from a colleague, vendor software, or AI. This is why “AI everywhere” is the wrong default for actuarial modelers. The question is not whether AI is good or bad. It’s where does AI belong.

Designing an Actuarial AI Operating Model

One approach is to separate actuarial work into zones. In an Autopilot zone, AI can be used freely for low-risk, structured tasks such as drafting document skeletons, summarizing meeting notes, formatting change logs, or creating basic test scaffolding. In a Co-pilot zone, AI can assist with close human review, as in prototyping calculation logic, translating requirements into pseudocode, or suggesting edge cases. In a Navigator zone, however, AI should function only as a cross-check, as a secondary sounding board to validate reasoning, not as the source of authoritative output. That includes assumption setting, governance decisions, and ambiguous debugging. The kind of tasks where correctness depends on context and judgment rather than pattern completion.

Instead of asking whether AI should be adopted, actuarial teams can ask where is validation easy, where is the blast radius low, and where is professional judgment irreducible. That is a more useful question and leads to better governance.

Then actuarial leaders should measure AI on real tasks rather than rely on impressions since felt speed can be misleading. Teams should test AI on actual actuarial work and track not just time saved at the front end, but rework, review burden, and error rates across the full cycle. They should define acceptable use cases, preserve documentation and auditability, and make sure that junior staff still develop the judgment that comes from doing hard work rather than merely supervising outputs.

That last point may be the most important. The first version of the AI productivity glitch was about speed. But the deeper issue may not be speed at all. Even if AI begins to accelerate some expert workflows, the more important question is whether it increases value or simply increases output. A profession can become more efficient and still become less capable of understanding the systems it relies on. In that case, the real risk is no longer just slowdown. It is the weakening of human judgment precisely when stronger judgment is becoming more necessary.

That is why the core lesson should remain steady. In expert, context-rich environments, speed is not the same as progress. Actuarial modeling intensifies that distinction because its outputs are shaped by tacit knowledge, vulnerable to silent error, and carried forward under professional accountability. AI can absolutely improve actuarial work. But it will do so only when it is used with discipline, measured against real outcomes, and kept firmly inside a framework where judgment stays human.

In the end, the profession's advantage will not come from using AI everywhere. It will come from knowing where AI genuinely helps, where it quietly harms, and where human rigor must remain non-negotiable.

Shane Leib, FSA, MAAA is Associate Teaching Professor at the University of Notre Dame and Principal of Irish Actuarial Consulting, LLC.



ACTUARIAL INTELLIGENCE BULLETIN

Case Study: GenAI-Driven Market Comparison

SIMON HATZESBERGER CERA, PHD AND IRIS NONNEMAN MSC AAG

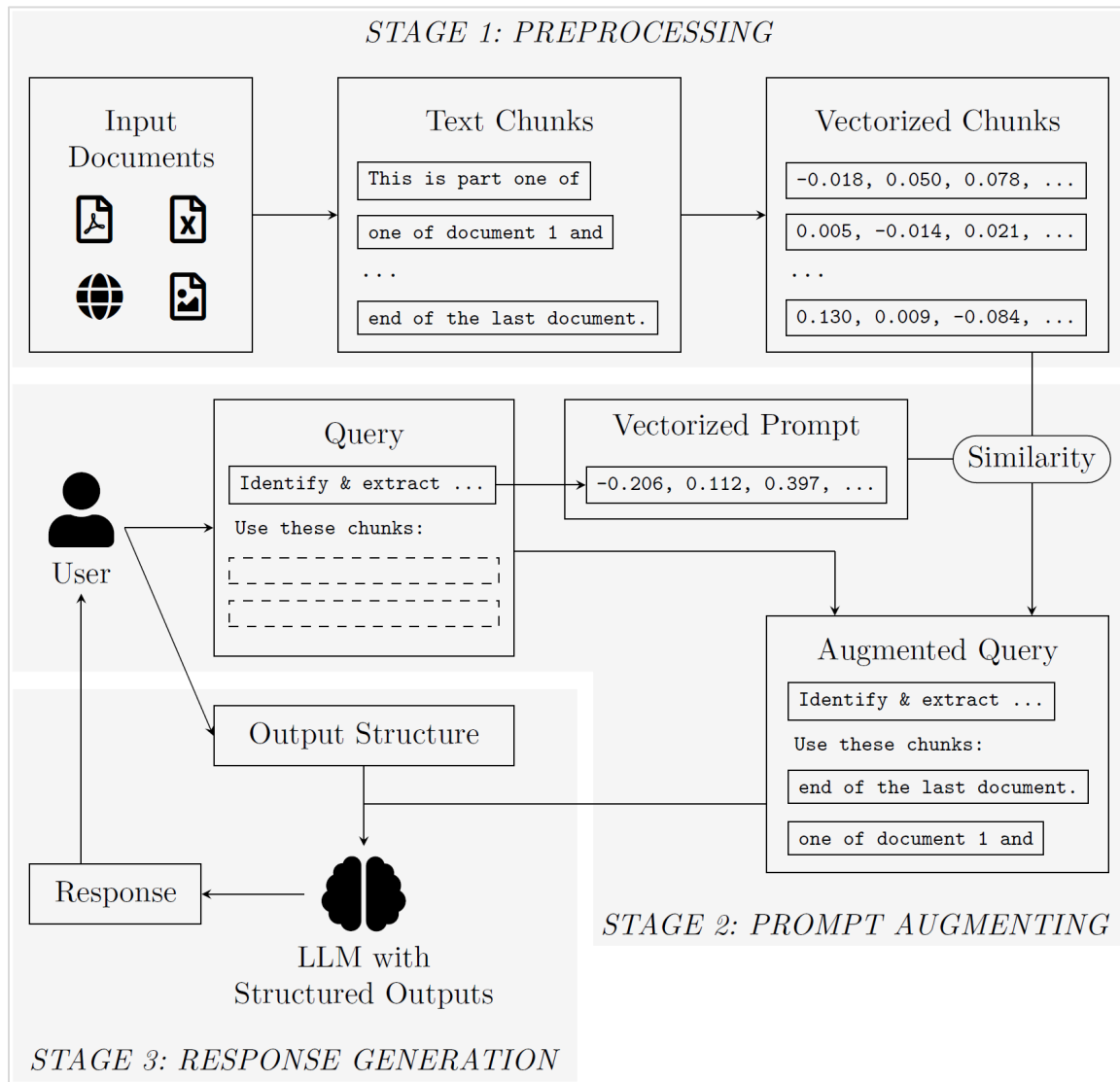
Extracting and harmonizing comparable financial and insurance data from annual reports is typically labor-intensive and error-prone: report structures are non-standardized, content formats vary widely, and information is scattered across lengthy documents. We demonstrate how generative AI can streamline this task by extracting three key aspects from the 2025 annual reports of three major European insurers (AXA, Generali, and Zurich): (1) regulatory capital ratios under Solvency II or the Swiss Solvency Test, (2) EUR discount rates for insurance contract liabilities by duration, and (3) insurer financial strength ratings from rating agencies.

To assess the robustness of the extraction pipeline, we benchmark five LLMs—Claude Sonnet 4.6, GPT-4.1, GPT-4.1 mini, GPT-5.4, and GPT-5.4 mini—under a shared RAG pipeline, with deterministic evaluation against manually extracted ground-truth values and repeated runs to quantify stability.

Approach and Techniques

The 3-stage pipeline (Figure 1) comprises Preprocessing, Prompt Augmenting, and Response Generation. For a fair model comparison, the retrieval pipeline is shared and held constant across all models; only the generation model in Stage 3 varies.

Figure 1
THE 3-STAGE RAG PIPELINE



Preprocessing converts input documents to plain-text chunks and embeddings. Prompt Augmenting retrieves relevant chunks by vector similarity. Response Generation uses an LLM with Structured Outputs to produce a response that adheres precisely to the user-specified schema. (Source: Authors)

Stage 1: Preprocessing

PDFs are converted to plain text, cleansed to remove formatting artifacts, and segmented via a sliding window into chunks of at most 2,000 characters with 300-character overlap. Each chunk is embedded using OpenAI's text-embedding-3-large model, producing numerical vectors that support semantic similarity retrieval.

Stage 2: Prompt Augmenting

For each extraction aspect, a task-specific prompt is embedded with the same model and compared against chunk vectors by cosine similarity. The top 10 chunks exceeding a 0.30 similarity threshold are concatenated with the

prompt to form the augmented query. Prompt specificity is critical: the ratings prompt, for instance, must distinguish insurer financial strength ratings from issuer credit or senior debt ratings.

Stage 3: Response Generation

The augmented query is sent to the LLM together with a Pedantic schema that defines the expected output (integers, floats, literals, lists). Structured Outputs ensure machine-readable responses that can be validated without free-text parsing. A LangChain wrapper provides a unified interface so that identical extraction logic can be applied to all five benchmarked models. The base pipeline can also be extended with advanced RAG variants—GraphRAG incorporates a knowledge graph to support multi-hop reasoning across related entities, while Agentic RAG embeds AI agents that dynamically select tools and refine retrieval strategies based on query complexity.

Results

All experiments use a temperature of 0, which minimizes randomness in LLM outputs and yields near-deterministic responses. Each model–aspect–company combination is executed 20 times, yielding 900 extractions in total. Ground-truth reference values were manually extracted by the authors from the 2025 annual reports, comprising 45 individual fields: 3 solvency ratios, 33 discount-rate points (10 for AXA, 18 for Generali, 5 for Zurich), and 9 ratings. Each extraction is scored 1/0: exact match for solvency and discount rates; alias-aware set comparison for ratings (e.g., "S&P Global Ratings" and "Standard & Poor's" both normalize to "S&P").

Table 1

PASS RATES (%) BY MODEL AND EXTRACTION ASPECT, AVERAGED ACROSS 20 RUNS AND THREE INSURANCE GROUPS

Model	Solvency Ratios	Discount Rates	Fin. Strength Ratings
Claude Sonnet 4.6	100.0	100.0	98.3
GPT-4.1	100.0	100.0	100.0
GPT-4.1 mini	100.0	100.0	66.7
GPT-5.4	100.0	100.0	100.0
GPT-5.4 mini	100.0	100.0	91.7

Source: AXA, Generali, Zurich.

GPT-4.1 and GPT-5.4 achieve perfect pass rates across all nine company–aspect combinations. Claude Sonnet 4.6 reaches 98.3% on ratings; GPT-5.4 mini reaches 91.7%; GPT-4.1 mini trails at 66.7%. Solvency-ratio and discount-rate extraction prove universally reliable across all five models.

Failure analysis localizes the rating errors to the generation stage, not retrieval: the relevant context is correctly fetched, but smaller models inconsistently resolve entity-level ambiguity (for example, AXA SA the holding company versus its primary insurance subsidiaries). This aligns with the broader literature, which distinguishes retrieval failures (insufficient context) from generation failures (incorrect reasoning over correct context). Retrieval failures are typically fixed by refining prompts or retrieval parameters; generation failures require more explicit instructions or a more capable model.

Implications for Actuarial Practice

RAG combined with Structured Outputs enables actuaries to obtain structured, comparable insights from lengthy technical reports with greater efficiency and consistency than manual extraction. The multi-model benchmark reveals that retrieval quality—chunking, embedding model, prompt design—matters as much as the choice of generation model; improvements to retrieval often yield larger gains than switching to a more capable LLM.

The approach is most effective on well-defined, structured aspects. Less structured areas require iterative refinement and tailored prompt engineering. When the desired information is absent, the model should be instructed to say so rather than hallucinate an answer. Human oversight remains strongly recommended: interim outputs such as retrieved chunks and comparisons with prior values support validation and reinforce trust in AI-driven analyses.

References

Full article: <https://arxiv.org/abs/2506.18942>

Notebook: https://github.com/IAA-AITF/Actuarial-AI-Case-Studies/tree/main/case-studies/2025/GenAI-driven_market_comparison

Simon Hatzesberger CERA, PhD is Manager Actuarial and Insurance Services, Deloitte
Iris Nonneman MSc AAG is Senior Data Scientist at Achmea



ACTUARIAL INTELLIGENCE BULLETIN

AI as a Second Opinion in Claims Reserving

YULIA NECHAY AND BEN ZICKEL

In non-life and health claims reserving, the central challenge has never been only about numbers. It has always been about judgment.

Traditional actuarial work relies on experience, professional reasoning, and careful interpretation of incomplete and often unstable data. Two highly competent actuaries can analyse the same portfolio, apply sound assumptions, and arrive at materially different results. This is not a weakness of the profession. It reflects the inherent uncertainty of insurance liabilities and the many legitimate modelling choices available.

The question is whether machine learning can strengthen actuarial practice by making results more consistent, more objective, and easier to validate. We believe the answer is yes, if AI is used as a disciplined second opinion.

Reserving Is a Landscape of Choices

Reserving involves numerous decisions. Actuaries choose model structures, define triangles, select levels of aggregation, treat large claims, incorporate inflation, account for changes in portfolio mix, select priors, and determine tail assumptions. Each step is necessary. Yet together these choices create substantial degrees of freedom, and small modelling decisions can lead to large differences in outcomes.

Validation is also challenging. The true ultimate cost of claims is known only years later, and intermediate validation is complex and not well defined. As a result, most validation focuses on plausibility rather than predictive accuracy. We check whether assumptions are reasonable and whether outputs appear consistent with experience. While essential, this approach can still leave room for inconsistency across teams, companies, and time.

AI can help reduce that inconsistency by providing an independent benchmark derived directly from the data.

BUT only if done properly.

Requirements for AI in Actuarial Practice

Because reserving decisions carry significant financial and regulatory consequences, any AI used in this context must meet particularly high standards. To be useful as a second opinion, it must be transparent, robust, stable, reproducible, and compatible with governance and regulatory expectations. It should express results in terms that actuaries can interpret and evaluate.

Without these properties, such methods are unlikely to gain acceptance within the actuarial profession.

Challenges of Applying Machine Learning to Insurance Data

On the other side, insurance datasets present unique challenges for machine learning. Data is often limited and interdependent, development is long, recent periods are incomplete, and structural changes may reduce comparability across years.

Highly flexible models can easily overfit such data. Therefore, effective actuarial ML approaches must incorporate domain knowledge (Bayesian priors) to remain robust with limited information.

Structured probabilistic models that explicitly represent uncertainty are particularly well suited to this environment, as they can produce realistic forecasts even when data is sparse.

These models should be composed of interpretable components reflecting real insurance practice. It would include exposure effects, development patterns, calendar influences such as inflation, portfolio evolution, and residual behaviour. These components mirror the drivers that actuaries already consider in traditional analyses and use in professional judgment, making the models easier to understand.

Model Selection

Proper model selection must be based on objective predictive performance rather than subjective preference. This predictive performance can be evaluated using classic machine learning methods like cross-validation.

Because reserving problems involve predicting entire future blocks of development, validation approaches should reflect that reality by assessing how well models forecast unseen portions of the data.

Actuarial teams can continue their established reserving process while running an independent machine learning framework on the same data. If the two results are broadly consistent, the AI output increases confidence and may reduce review effort. If they differ materially, the discrepancy becomes valuable information rather than a failure.

Differences may reveal hidden assumptions, structural changes in the portfolio, data issues, or areas where expert judgment has a strong influence. In this way, AI becomes a data driven benchmark that stimulates discussion and deeper analysis.

Since probabilistic models produce full distributions, their quality can be evaluated using scoring approaches that reward accurate forecasts while penalizing both bias and excessive uncertainty. Systematic deviations may indicate overconfidence, underestimation of variability, or structural bias.

Why the “Second Opinion” Approach Works

Introducing AI as a parallel analysis offers a practical and low risk pathway for adoption.

Actuarial teams can continue their established reserving process while running an independent machine learning framework on the same data. If the two results are broadly consistent, the AI output increases confidence and may reduce review effort. If they differ materially, the discrepancy becomes valuable information rather than a failure.

Differences may reveal hidden assumptions, structural changes in the portfolio, data issues, or areas where expert judgment has a strong influence. In this way, AI becomes a data driven benchmark that stimulates discussion and deeper analysis.

This approach respects professional expertise while improving objectivity. It also aligns with governance principles that favour independent model validation.

Implications for the Role of the Actuary

Used appropriately, AI does not diminish the role of actuaries. It changes it.

Machine learning can automate the production of structured forecasts and uncertainty estimates. Actuaries remain essential for interpreting results, incorporating contextual knowledge, evaluating assumptions, and ensuring that decisions remain transparent and defensible.

In this sense, the actuarial role shifts from producing numbers to ensuring the reasoning behind them is clear and defensible. This evolution is consistent with the broader movement toward model governance and enterprise risk management.

A Path Forward

The future of reserving should not be framed as a choice between actuarial judgment and machine learning. The strongest framework combines both.

Human expertise provides context, experience, and accountability. Explainable probabilistic models provide consistency, objectivity, and measurable predictive performance. Together they can produce results that are more robust than either approach alone.

Using AI as a second opinion offers a pragmatic entry point. It delivers value immediately while building trust over time. We don't believe that artificial intelligence will replace the actuary. We think it will help actuaries make more informed, transparent, and defensible decisions in the face of uncertainty.

Yulia Nechay is Chief Actuary and Product Strategy, PredicTri and Ben Zickel is CTO, PredicTri.

Map, Measure, Govern, and Manage: An Actuarial View of the Financial Services AI Risk Management Framework

VAJIRA MANATHUNGA, PH.D., ASA

This is the first of two articles that discuss material from the Financial Services AI Risk Management Framework (FS AI RMF) which is a specialized, industry-led adaptation of the NIST AI Risk Management Framework (NIST AI 100-1), specifically tailored for the financial sector. Ed.

Many view AI governance as a new administrative hurdle, but for actuaries, the 230 control objectives of the new FS AI RMF represent something much more familiar in the professional standards we already live by. Actuaries have long been in the business of quantifying uncertainty. As artificial intelligence increasingly influences pricing, underwriting, reserving, fraud detection, customer service, portfolio management, and other operations across the financial sector, actuaries face a new challenge: how do we map, measure, govern, and manage the risks of the AI models we rely on to measure risk?

In February 2026, the Cyber Risk Institute (CRI), working with over 100 financial institutions, the Financial Services Sector Coordinating Council (FSSCC), and the U.S. Department of the Treasury, published the Financial Services AI Risk Management Framework (FS AI RMF). Structurally aligned with the National Institute of Standards and Technology (NIST) AI Risk Management Framework, it provides 230 Control Objectives designed to help financial organizations manage AI risk in a structured and scalable way.

The framework organizes AI risk management into four core activities: Govern, Map, Measure, and Manage. While these labels come from the AI risk management world, the underlying logic is familiar to actuaries. Actuaries are trained to think in terms of the Control Cycle: define the problem, design the solution, monitor results, and respond as conditions change. In many ways, the framework aligns naturally with both the SOA Control Cycle and the broader principles embedded in the Actuarial Standards of Practice (ASOPs). For actuaries, this connection matters. AI governance should not be seen as something external to actuarial work rather than extension of same disciplined thinking actuaries already possessed.

Govern

The Govern function is the bedrock of the framework. It focuses on the organizational foundation needed for trustworthy AI, including policies, accountability structures, training, oversight, documentation, stakeholder engagement, and third-party risk processes. Govern is the largest of the four functions, encompassing 81 of the framework's 230 control objectives. It asks questions such as: Who is accountable for AI risk? Are there policies defining acceptable AI use? Are legal, compliance, risk, and business leaders involved? Are there procedures for incidents and audits? As actuaries increasingly work with vendor-provided AI tools, Govern places particular emphasis on third-party risk management. If governance is weak, even technically strong AI systems may introduce unacceptable legal, operational, reputational, or ethical risks. In Control Cycle terms, this is the context-setting that precedes any problem definition. In ASOP terms, ASOP 58, Enterprise Risk Management, is the strongest parallel to Govern because it covers governance over risk processes, risk identification, risk classification, risk appetite, and risk inventories. ASOP 41, Actuarial Communications, also connects closely through its emphasis on transparency and disclosure. In addition, ASOP 56, Modeling, requires actuaries to ensure that models operate within an appropriate governance framework, including structures designed to reduce model risk.

Map

The Map function, with 47 control objectives, requires organizations to identify and document the context, assumptions, limitations, and potential impacts of AI systems before deployment. Before an AI system can be

trusted, the organization must understand what the system is intended to do, what business objective it supports, what data it uses, who is affected, and what the consequences of error may be. Actuaries will recognize this as the first stage of the Control Cycle: define the problem. ASOP 56 again provides a strong connection because it requires the actuary to understand the model, its capabilities, limitations, and boundaries. That is precisely what Map demands for every AI system.

Measure

Once the AI system and its context are understood, the next step is to measure how well it performs and what risks it creates. The Measure function provides this through 59 control objectives, translating identified risks into metrics, tests, and ongoing evaluations across the full spectrum of AI trustworthiness: validity, reliability, safety, security, transparency, explainability, privacy, fairness, and resource efficiency. This broader view is important. An AI system can perform well on average and still create serious problems. It may work accurately overall but fail for certain subgroups. It may be stable in testing but unstable in production. It may improve efficiency while introducing opacity or compliance risk. Actuaries are already familiar with the idea that no single metric tells the whole story. AI requires a similar mindset. The clearest ASOP parallels here are ASOP 23, ASOP 25, and ASOP 56.

Manage

The Manage function, with 43 control objectives, corresponds to the “monitor the results” and “respond to experience” stages of the Control Cycle. It closes the loop by prioritizing risk responses, maximizing benefits, and ensuring ongoing operational resilience. It is where assessment becomes action. Manage requires formal go/no-go criteria before deployment, clear strategies for risk mitigation, transfer, avoidance or acceptance, and ongoing post-deployment monitoring. Sometimes the right response is to proceed with controls, impose additional oversight, restrict the use case, or reject deployment altogether. Residual risks must be documented and communicated to downstream stakeholders, paralleling the actuarial obligation under ASOP 41 to provide clear disclosures about limitations in work product.

Although the four functions are presented separately, their real strength lies in how they work together. Govern sets expectations and accountability. Map clarifies the system and its context. Measure evaluates performance and risk. Manage drives action and improvement. Together, they form a comprehensive approach to responsible AI adoption. For actuaries, the deeper parallel between the FS AI RMF and the Control Cycle is philosophical. Both reject the idea that risk management is a one-time deliverable. Both insist on continuous feedback loops: governance informs how risks are mapped; mapped risks drive what is measured; measurement results shape how risks are managed; and management outcomes feed back into governance.

As AI continues to influence financial services, actuaries have an opportunity to contribute not only as users of these tools, but also as stewards of sound practice. The profession’s grounding in the Control Cycle and the ASOPs makes actuaries especially well-suited to bring rigor, discipline, and accountability to AI risk management. The FS AI RMF is not a replacement for actuarial standards, but a specialized toolkit that allows us to apply long-standing actuarial principles to the complexities of the AI age.

Vajira Manathunga, Ph.D., ASA is Director of Actuarial Science Program and Associate Professor, Department of Mathematical Sciences, Middle Tennessee State University.

Stages of AI Adoption for Businesses

ERICK MWANGI MUGUCHIA

This is the second of two articles that discuss material from the Financial Services AI Risk Management Framework (FS AI RMF) which is a specialized, industry-led adaptation of the NIST AI Risk Management Framework (NIST AI 100-1), specifically tailored for the financial sector. Ed.

For financial institutions and risk-driven industries, understanding the maturity of AI adoption is critical for balancing innovation with control.

The Financial Services Artificial Intelligence Risk Management Framework (version 1.0) released in February 2026 by the Cyber Risk Institute (CRI) details four stages of AI adoption of businesses, highlighting the capabilities, risks and governance requirements at each stage.

Stage 1: Initial

At this stage, AI is new and largely experimental. Organizations explore proof-of-concepts, isolated research projects and pilot studies to assess whether AI can make a meaningful impact. Integration with business processes is very minimal and governance is virtually absent.

Key features:

- AI is treated as explanatory research.
- Minimal deployment, no formal governance.
- Focus on learning and assessing potential impact.

As organizations move beyond experimentation, AI begins to enter a real operational environment.

Stage 2: Minimal

Here, one or a few AI systems are deployed into production, but they remain isolated from core business systems. AI may supplement workflows, for example, fraud scoring or document classification, but it is not yet embedded in broader actuarial or underwriting infrastructure.

Key features:

- Limited deployment, isolated tools.
- Here the company starts to experience early exposure to AI risks.
- Governance frameworks are still immature.

With increased reliance on AI, complexity and risk begin to scale significantly.

Stage 3: Evolving

This stage marks a qualitative shift. At this stage, AI risk transitions from isolated incidents to enterprise-wide exposure. AI systems are deployed across multiple business functions and increasingly integrated with core systems. In financial services, this includes, underwriting, claims, fraud detection, customer service and actuarial models.

Organizations recognize the need for dedicated governance infrastructure. Model risk management frameworks are extended to AI and independent validation begins. Cross-functional teams such as actuaries, data scientists, engineers and risk professionals collaborate to manage complexity.

Key risks:

- Rapid growth in production models outpaces governance capacity.
- Discriminatory outcomes in underwriting and pricing create regulatory exposure. For example, an AI-driven underwriting model may unintentionally introduce bias if trained on historical imbalance data, exposing the organization to regulatory and reputational risk.
- Explainability becomes critical across thousands of daily decisions.
- AI-driven workflows risk brittleness when models fail or change.

Minimum governance requirements:

- Independent validation with AI expertise.
- Automated monitoring and incident response.
- Explainability tools (e.g. SHAP) for customer-facing decisions.
- Formal bias testing programs.
- Strong data governance with lineage and retention policies.
- Defined change management processes.

At full maturity, AI becomes inseparable from the organization's core strategy.

Stage 4: Embedded

At this stage, AI is woven into the operational and strategic fabric of the organization. Decisions once made by humans using AI as a tool are now executed by AI systems under governance frameworks. Products, pricing, customer interactions, capital allocation and regulatory submissions are substantially shaped by AI.

AI becomes a source of competitive advantage but also a source of systemic risk. Failures are not just operational incidents but strategic events.

Examples of applications:

- Fully autonomous underwriting for standard risks.
- Behavioral AI optimizing distributions and renewals.
- Enterprise risk AI integrating underwriting, reserving, investments and capital.

Governance requirements:

- AI Managements Systems (AIMS), analogous to ISO 27001 for information security.
- Real-time data quality monitoring.
- Privacy-enhancing technologies.
- Board-level accountability for AI as an institutional asset class.

Conclusion

Understanding these stages is strategically essential. Progression is not just a technical upgrade, it requires parallel advancement in governance, organizational capability, data infrastructure and regulatory posture.

Businesses that master these transitions position themselves for sustainable competitive advantage, while those that lag risk exposure to financial, regulatory and reputational consequences. The future of business will not simply use AI; it will be determined by how well AI is governed.

For actuaries and risk professionals, mastering AI adoption is not optional, it is central to safeguarding financial stability and ensuring fairness in a data-driven future.

Erick Mwangi Muguchia, Actuarial Student, BSc Actuarial Science (Expected 2026), Meru University of Science and Technology

ACTUARIAL INTELLIGENCE BULLETIN

From Drafts to Done: Using AI to Streamline Communication

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 summarization and communication.

In a technical field like actuarial work, the ability to communicate effectively, especially to senior leaders, can set one employee apart from another. In a poll of over 500 actuaries published on ProActuary, the top answer to the question, “What made you more valuable as an actuary?” was communication skills.¹ A 2024 list of most in demand skills published on LinkedIn indicates that communication is the number one skill employers look for.² A 2024 Grammarly study showed that knowledge workers—those who work with data—spend about half their time on written communication.³ To optimize time, and value, created through written communication, generative AI can help.

Before exploring how generative AI can help optimize written communication, here are three email best practices.

- 1) In my article, “Business Writing Insights: Less is More?” I referenced the importance of the “scroll down rule.”⁴ This has to do with keeping communications, especially emails, to one page. This increases the likelihood that a recipient will read and respond faster. That equates to about two hundred words in Microsoft Outlook.
- 2) I referenced the significance of writing at an 8-10 Flesh Kincaid grade equivalent level readability score. Per Inc.com, the average person receives, a whopping 121 emails a day!⁵ It is necessary to send communications that are easy to understand quickly.
- 3) I wrote about getting to the point and leading with the recommendation. This leaves no doubt about the main message and intended purpose.

As a subject matter, I am going to use a ~2,000 word article I wrote in The Actuary magazine about AI ethical principles, standards, and best practices.⁶ I am going to use AI to draft an email to senior management about reviewing the AI governance program. I will use two large language models (LLMs), Claude Sonnet 4.5, and Chat GPT-5-MINI, for this exercise.

¹ “Actuarial Communication Skills,” *ProActuary*, <https://proactuary.com/actuarial-communication-skills/>, Accessed April 28, 2026.

² LinkedIn, “The Most In-Demand Hard and Soft Skills,” *LinkedIn Talent Blog*, <https://www.linkedin.com/business/talent/blog/talent-strategy/linkedin-most-in-demand-hard-and-soft-skills>, Accessed April 28, 2026.

³ “Time Management Statistics,” *Clockify*, 2026 edition, <https://clockify.me/time-management-statistics>, Accessed April 28, 2026.

⁴ Mitchell Stephenson, “Business Writing Insights: Less is More?,” *The Actuary*, September 2025, <https://www.theactuarymagazine.org/business-writing-insights-less-is-more/>.

⁵ Joy Gendusa, “People Receive 121 Emails a Day,” *Inc.*, March 17, 2025, <https://www.inc.com/joy-gendusa/people-receive-121-emails-a-day/91162191>.

⁶ Mitchell Stephenson, “All Things Artificial Intelligence,” *The Actuary*, July 2024, <https://www.theactuarymagazine.org/all-things-artificial-intelligence/>.

The Prompt

Write a one-page email for senior management which makes a recommendation about reviewing their current AI governance program. Based the recommendation on the content in the article “All Things Artificial Intelligence” in the July 2024 edition of *The Actuary* publication. Keep it to no more than two hundred words. Write it at a Flesh-Kincaid grade equivalent readability score of between 8-10. State the key conclusion first and then list supporting evidence.

Chat GPT Results

Dear Senior Management Team,

Key conclusion: I recommend we conduct a focused review and refresh of our current AI governance program within the next quarter to ensure it remains fit for purpose as AI use expands across the organization.

Supporting evidence from “All Things Artificial Intelligence” (The Actuary, July 2024):

- AI adoption is accelerating faster than most governance frameworks were designed to handle, increasing model, operational, and reputational risk.
- Effective AI oversight requires clear accountability, including defined roles for management, model owners, and independent review.
- Traditional model risk controls are necessary but not sufficient for AI; added attention is needed for data quality, bias, explainability, and ongoing monitoring.
- Governance should cover the full AI lifecycle, from design and procurement through deployment, use, and retirement.
- Regular training for leaders and staff is critical so decisions about AI use are informed, consistent, and aligned with professional standards.

A short, structured review would allow us to benchmark our practices against emerging expectations, identify gaps early, and reinforce confidence with regulators, clients, and employees.

Claude Results

Dear Senior Management,

I recommend we conduct an immediate review of our AI governance program to address emerging risks and opportunities.

Supporting evidence from recent industry analysis includes:

- AI tools can make errors and invent facts. Legal firms have submitted briefs with fabricated cases, and major tech companies have demonstrated inaccurate AI outputs in promotional materials.
- The NAIC Model Bulletin recommends specific governance elements: committee structure, senior management ownership of AI strategy, internal audit reviews, and evidence retention for compliance.
- Industry growth is substantial. AI's value in insurance is projected to reach \$80 billion by 2032, up from \$4.5 billion in 2022.
- Best practices include templates for output analysis, periodic attestations that results match intended purposes, and documented rollback plans.
- Actuarial Standards of Practice (ASOPs 12, 54, and 56) provide guidance on model reliability, validation, and sensitivity testing that should inform our governance framework.

Given these risks and opportunities, we should verify our current program addresses these evolving requirements.

Observations

Both LLMs do an excellent job summarizing the content. The supporting bullets are quite different, though both valid. They use less than two hundred words, to fit on one page. Both lead with the conclusion and follow it with evidence from the article. Both LLMs fail to meet the targeted grade equivalent level, coming in around 15-16. I will need further prompts to improve readability scores.

It took less than one minute each to generate the content. As compared to the effort to perform research using AI I documented in my March 2026 article: Using AI to do Research about AI, this exercise was incredibly straightforward and returns desired content on the first prompt from both LLMs, excepting the readability scores.

Takeaways

A commonly articulated ethical principle is accountability.⁷ It is important to understand each bullet, the conclusion, and to stand behind it. It is imperative to read and review the content before sending, checking for grammar, misspellings, words, or phrases out of context, and content. Taking these steps will add an essential element of accountability.

It is no wonder that technical professionals like actuaries rely on generative AI for summarization and communication as a primary use. This use will continue to add value to individuals looking to optimize their time spent on written communication and to improve their value to their employers.

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.



**ACTUARIAL
INTELLIGENCE
BULLETIN**

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

Building Fair and Accurate Predictive Models: An Interdisciplinary Imperative for Modern Insurance

ALEXANDRE BOUMEZOUED, PHD, FIA; FRANÇOIS HU, PHD; MICHAEL NIEMERG, FSA, MAAA; BERTILLE TIERNY, FIA

Ensuring fairness in predictive models is a critical component of any successful AI initiative. Implementing fairness testing is complex and can be driven by regulatory expectations, reputational considerations, and operational integrity. Furthermore, fairness in insurance is not solely technical: it is an interdisciplinary problem that requires judgment from domain experts, compliance teams, and AI engineers. Failing to integrate these perspectives produces blind spots that technical fixes alone cannot resolve.

An Interdisciplinary Approach to Fairness Assessment and Remediation

The interdisciplinary nature of fairness is essential. Underwriters provide business and operational context that grounds model design in how decisions are made, help embed AI into real-world business processes, and can provide real-time feedback and auditing as AI is operationalized. Compliance teams interpret regulatory constraints, assess the permissibility of variables, and ensure audit-ready documentation. AI engineers and data scientists implement testing methods, detect proxy effects, and design and apply quantitative mitigation strategies. Actuaries can be part of those developments, and evaluate their impacts on pricing, risk selection, and longer-term portfolio behavior.

Without this interdisciplinary perspective, fairness work can fail in predictable ways: models may be calibrated to the wrong underwriting and business context, leading to both incorrect and unfair rating assignments, and mitigation choices may be operationally infeasible or noncompliant, creating audit and reputational risk even when statistical metrics look strong.

Setting Up the Challenge

Bias can be defined in multiple ways and can enter a modeling pipeline through several mechanisms. Sensitive attributes can influence predictor variables directly, and complex models may combine multiple features in ways that unintentionally create group-level disparities. Even when sensitive attributes are removed, correlations in the data may lead the model to infer proxies and reproduce potentially unequal outcomes. Understanding these mechanisms is essential to diagnosing whether differences in model behavior across groups reflect legitimate risk variation or artifacts of data and model structure.

Choosing the Right Fairness Metric

Fairness in predictive modeling can be assessed e.g. using group fairness metrics. Demographic parity evaluates whether predicted outcomes differ between groups, without conditioning on the true outcome. Predictive parity assesses whether individuals receiving the same modeled decision experience similar true outcomes regardless of group membership. Each metric highlights different aspects of fairness, and no single measure captures all facets of model behavior. Selecting an appropriate metric requires understanding the business use case and the decision processes that the model supports.

Bias Mitigation and the Fairness-Accuracy Trade-Off

Successful bias mitigation interventions can occur at different stages of the modeling pipeline. Pre-processing modifies the data to reduce structural imbalances. In-processing incorporates fairness constraints into the model's objective function. Post-processing adjusts the model's outputs and can therefore be implemented without retraining. Each approach offers different advantages depending on data availability, model type, and operational constraints.

When thinking about these bias mitigations, make sure you have a good understanding of the use case and regulatory considerations. For instance, many types of bias mitigation techniques require using race explicitly to adjust the model’s predictions, but this may not be operationally feasible or may not be something regulators would accept.

A common concern is that fairness may come at the expense of model performance. This trade-off can be managed: some empirical studies across insurance applications show that models can achieve substantial reductions in group disparities with only marginal reductions in predictive power. To illustrate how mitigation works in practice, we turn to a concrete case in mortality risk scoring.⁸

Fairness in Practice: Mortality Risk Scoring

To illustrate how post-processing mitigation can be applied in a realistic setting, we consider mortality risk scoring using the GOSSIS⁹ dataset (more than 130,000 intensive care unit visits over one year across multiple countries and 200+ U.S. hospitals). The dataset provides a black-box mortality risk score aligned with the APACHE IV probabilistic prediction of in-hospital mortality. We choose Predictive Parity as the reference fairness metric, and we stratify the dataset by sensitive group using ethnicity as a proxy.

Choosing the fairness criterion and tolerance remains a judgment call that benefits from joint actuarial, compliance and engineering oversight and governance throughout the model lifecycle.

To stress-test the framework, we introduce an artificial group shift by adding 0.1 to all scores for one sensitive group and subtracting 0.1 from the other. On this biased variant, the black-box score becomes significantly unfair for Predictive Parity, meaning mortality predictions within actual mortality cases differ materially between ethnic groups. Applying the Lagrange-based post-processing method, which adjusts the model’s output decision thresholds via an optimization procedure to better satisfy fairness constraints, reduces Predictive Parity unfairness from 0.199 to 0.017 while maintaining overall predictive quality.

This example shows that controlled bias testing and mitigation can reveal how easily group shifts create inequities, and how post-processing methods can correct them. Choosing the fairness criterion and tolerance remains a judgment call that benefits from joint actuarial, compliance, and engineering oversight and governance throughout the model lifecycle.

About Governance

Model governance structures benefit from integrating fairness considerations throughout the model lifecycle. Clear fairness objectives can be established before model development or enhancement begins. Variable selection, feature engineering, and model calibration can be reviewed through the lenses of both predictive performance and fairness. Continuous monitoring is essential, as model behavior can evolve over time with changes in business mix, data quality, and operating conditions. Documentation enhances transparency and supports internal and external review of how fairness is assessed and maintained.

⁸ François Hu and Alexandre Boumezoued, “Trustworthy Artificial Intelligence in Insurance: Navigating Fairness and Performance in Predictive Modeling,” *Milliman Insight*, June 19, 2025, <https://www.milliman.com/en/insight/trustworthy-ai-insurance-fairness-predictive-modeling>.

⁹ Kaggle, “1056Lab Patient Survival Prediction,” *Kaggle Competitions*, <https://www.kaggle.com/competitions/1056lab-patient-survival-prediction>.

Fairness as Part of the Operating Model

Fairness in predictive modeling reflects a broader shift in how insurance organizations think about model governance. As more sophisticated models become integrated into core processes, fairness becomes a measurable attribute that must be monitored and managed like any other risk dimension. Interdisciplinary collaboration is therefore not optional; it is the operating model required to ensure that predictive systems remain accurate, lawful, and consistent with the industry's obligations to customers.

The right approach is to operationalize fairness: standardize it, set objectives up front, document assumptions and decisions, monitor performance and disparities over time, and convene underwriting, compliance, actuarial, and engineering early to agree on metrics, tolerances, and remediation triggers. This type of standardized, interdisciplinary approach will ensure you can simultaneously ensure fairness while still hitting all your business objectives.

Alexandre Boumezoued, PhD, FIA, Principal and AI Director; François Hu, PhD, Lead AI Engineer; Michael Niemerg, FSA, MAAA, Principal and Director, Data Science and Analytics; Bertille Tierny, FIA, AI Engineer at Milliman.



ACTUARIAL INTELLIGENCE BULLETIN

NIST Launches Trustworthy AI Initiative for Critical Infrastructure

DAVID INGRAM, FSA, CERA

On first thought, the currently popular forms of AI, like chatbots and generative AI, do not seem to be a good match with the operation of critical infrastructure. Hallucinations and unpredictable model drift are not what you want in a power grid. These systems demand stability, predictability, and clear cause-and-effect relationships. The AI that we hear so much about right now does not naturally offer those qualities.

The Challenge of Unpredictability

The concern is not just hallucination in the narrow sense. It is the broader issue that the model does not “know when it doesn’t know,” and it does not operate within hard boundaries unless those are imposed externally. In a critical infrastructure context, the real risk is not that the AI fails frequently. It is that it fails rarely, but in ways that are not predictable or contained. That is exactly the kind of risk profile that operators of these systems work very hard to eliminate.

At the same time, the direction of travel is clear. Modern AI is already being embedded into these environments, initially in supporting roles such as monitoring, anomaly detection, optimization, and cybersecurity response. Those roles are expanding because they provide real value. Over time, the boundary between support and control begins to blur, particularly in situations where response times matter and human intervention becomes a bottleneck.

This is where the concept of trustworthiness becomes central. Trustworthiness is not a claim about an AI system being reliable in the abstract. It is a requirement that the entire system—of which the AI is only one component—behaves in a way that is predictable, observable, and controllable under both normal and stressed conditions.

From Model Intelligence to System Discipline

The key shift is from focusing on the intelligence of the model to focusing on the behavior of the system. A raw AI model is not something you would trust in a critical environment. A system that uses AI within carefully defined constraints may be.

In practice, this means that the AI system is not given direct authority over critical actions. It proposes, flags, or recommends. Its outputs are filtered through guardrails, deterministic logic, operational rules, or human oversight before anything irreversible occurs. Hybrid architectures, which combine the power of modern AI techniques with more classical AI approaches, some of which do offer greater stability, predictability, and explainability, can help to bring the best of the old and the new. Systems can be designed so that when the AI is wrong—and it will be wrong—the consequences are limited, visible, and recoverable.

This is the direction being articulated by the National Institute of Standards and Technology (NIST) in its work on trustworthy AI for critical infrastructure. The focus is not on making AI safe in isolation, but on defining how to build systems around AI that enforce properties such as fail-safe behavior, rigorous validation, and resistance to adversarial manipulation. In effect, the problem is reframed from one of model performance to one of system discipline.

NIST has launched an AI Risk Management Framework (AI RMF) Profile specifically focused on Trustworthy AI in Critical Infrastructure. The intent is to take the existing AI RMF and contextualize it into concrete, operational guidance for environments where failure has physical and systemic consequences. That includes Information Technology (IT), Operational Technology (OT), and Industrial Control Systems (ICS), where AI is increasingly being integrated into monitoring, control, and response functions.

The profile emphasizes a set of properties for highly risk-sensitive environments. Some situations call for deterministic behavior and fail-over to safe states when a problem is detected. They must be robust against adversarial manipulation across the full lifecycle. They must undergo rigorous testing, evaluation, validation, and verification (TEVV), not just at deployment but continuously as conditions change. And they must degrade gracefully, providing operators with clear, interpretable signals when performance is compromised.

Equally important, NIST is not positioning this as a closed standard. They are explicitly calling for participation from across the ecosystem—operators, vendors, regulators, policymakers, and researchers. The objective is to surface real-world use cases, identify gaps in current guidance, and align the framework with existing industry practices and constraints.

For practitioners, this is an opportunity to shape how trustworthiness is defined and operationalized in high-stakes environments. It is also a signal that the conversation is moving beyond whether AI will be used in critical infrastructure to how it will be governed. Those who engage early will have disproportionate influence over the standards that ultimately define acceptable practice.

Join the Movement

By working collaboratively, we can ensure that AI deployment in our nation's most essential sectors is backed by confidence and rigorous risk management.

How to get involved:

- **Sign up** for the mailing list to receive significant announcements.
- **Join** the Community of Interest and participate in informal discussions.
- **Download** the Concept Note to dive deeper into the project's technical goals.
- <https://www.nist.gov/programs-projects/concept-note-ai-rmf-profile-trustworthy-ai-critical-infrastructure>

Dave Ingram, FSA CERA is Managing Editor, Actuarial Intelligence Bulletin.



Insurance for AI Makers

KETAKI BORKAR & KIM HARRIS

The insurance market for Artificial Intelligence is in a rapid state of transition. Historically, AI-related claims have fallen under "silent AI" coverage. What this means is that these risks were absorbed by traditional insurance policies that didn't explicitly mention AI. The market is now moving, slowly but clearly, toward more explicit wording or "affirmative AI" coverage, introducing specific endorsements, exclusions, and standalone policies tailored to the unique risks of artificial intelligence.

The type of coverage a business needs depends heavily on whether they are building the technology (makers of AI) or are using it (users of AI). Given below is a breakdown of the different types of insurance available for both makers and users of AI.

This first article (first of a two-part series) focuses on organizations that build, train, license, or deploy AI products into the market.

1. Technology E&O (Tech Professional Liability)

Silent AI

The baseline coverage for makers. It covers financial losses suffered by customers due to the failure, negligence, or defective output of the AI technology.

Primary Objective: Protects the balance sheet against lawsuits alleging the AI failed to perform as promised.

Primary Coverage: Legal defense costs and damages for tech failures and negligent design.

Illustrative Claims Scenario: A commercial client deploys an AI document-classification engine sold by the maker. The model misroutes a large volume of claims files for weeks, causing missed SLAs, rework costs, and client penalties. The client sues the AI vendor for financial loss.

2. Media Liability

Silent AI

Covers content-based offenses generated by AI, such as defamation, libel, and inadvertent copyright/trademark infringement in output.

Primary Objective: Critical for Generative AI platforms that output text, images, or video that could legally expose the maker to "publishing" torts.

Primary Coverage: Defense and damages for defamation, privacy torts, and plagiarism.

Illustrative Claims Scenario: A user prompts a generative AI platform to create promotional copy, and the output falsely accuses a competitor of fraud. The affected party sues the platform maker alongside the user for defamation.

3. Cyber / Privacy Liability

Silent AI

Protects the maker from costs associated with data breaches, including prompt-injection attacks that leak proprietary training data or user PII.

Primary Objective: Paying for incident response (forensics, notification) and regulatory fines after a privacy event.

Primary Coverage: Forensics, ransom payments, third-party privacy liability, data restoration.

Illustrative Claims Scenario: A prompt-injection attack causes a multi-tenant AI application to expose confidential customer prompts and personally identifiable information. Multiple clients bring privacy claims and regulators request information.

4. Commercial General Liability (CGL) / Product Liability

Silent AI

Covers third-party bodily injury or physical property damage caused by AI (e.g., a warehouse robot injuring a worker).

Primary Objective: Essential for makers of "embodied AI" (robotics, autonomous hardware, medical devices).

Primary Coverage: Defense and settlements for physical injury lawsuits.

Illustrative Claims Scenario: An AI-enabled warehouse robot misidentifies a worker's position and collides with them, causing bodily injury. The injured worker brings a product liability claim against the manufacturer.

5. Standalone / "Affirmative AI" Liability & Warranty

Affirmative AI

Specialty coverage strictly for AI model underperformance against defined KPIs (accuracy, bias, uptime).

- Primary Objective: A sales enablement tool. It closes coverage gaps by guaranteeing the AI's efficacy to hesitant enterprise buyers.
- Primary Coverage: Financial remedy if the AI falls below agreed statistical performance thresholds.
- Illustrative Claims Scenarios: A fraud-detection model is sold with a contractual minimum precision threshold. Performance drops materially below the insured KPI and the customer seeks compensation under the warranty framework.
- A hiring-screening AI is represented as meeting a defined bias tolerance band. An audit later shows the model exceeds the agreed threshold, triggering a covered remediation or compensation mechanism.

6. D&O (Management Liability)

Silent AI

Protects the board and executives from shareholder lawsuits alleging mismanagement of AI strategy or misleading disclosures about AI capabilities ("AI Washing").

- Primary Objective: Critical for public companies or heavily funded startups facing investor scrutiny.
- Primary Coverage: Defense and settlements for securities class actions.
- Illustrative Claims Scenario: Executives are accused of "AI washing" after overstating how much of the company's projected revenue depends on proprietary AI capability. A securities action follows a sharp stock decline.

7. Standalone Intellectual Property (IP) Liability

Affirmative AI

Dedicated, high-limit coverage for patent infringement and copyright disputes regarding model training data (e.g., the New York Times v. OpenAI style risks).

- Primary Objective: Existential protection for Foundation Model developers scrapping the web for data.
- Primary Coverage: Legal defense and damages for infringement claims.
- Illustrative Claims Scenario: A publisher alleges that the maker trained its model on copyrighted content without authorization and seeks substantial damages and an injunction against further use.

8. Product / Technology Recall

Silent AI

Covers the first-party costs of pulling a defective AI product from the market or issuing a massive emergency patch to prevent imminent harm.

Primary Objective: Vital for automotive AI or medical AI where a "bug" endangers lives.

Primary Coverage: Notification costs, logistics, extra expense to recall/patch, reputation management.

Illustrative Claims Scenarios: An automotive supplier discovers that an AI-assisted driver-monitoring component intermittently fails to detect drowsiness, requiring recall or major remediation.

A medical-adjacent monitoring device with embedded AI generates dangerously unreliable alerts after a software update, forcing immediate customer notification and field remediation.

9. Transactional Liability / Reps & Warranties (R&W)

Silent AI

Covers financial loss arising from a breach of the seller's representations regarding their AI assets (IP, data rights, bias) during M&A.

Primary Objective: Protects an acquirer from discovering post-close that the AI startup they bought was built on stolen data or non-compliant code.

Primary Coverage: Defense and damages for breaches of reps related to IP ownership and data privacy compliance.

Illustrative Claims Scenario: After acquiring an AI startup, the buyer discovers that a key training dataset was licensed only for research and not for commercial deployment. The buyer seeks recovery under the insured representation package.

Conclusion

For AI makers, the real issue is not whether insurance exists, but whether the program is appropriate for how the AI is built, trained, sold, and supported. A model developer with copyrighted training-data exposure, an embedded-AI hardware manufacturer, and a SaaS vendor offering performance guarantees are not carrying the same risk, even if all three call themselves AI companies. The strongest insurance strategy for makers is therefore layered.

As the market matures, the winners will be the companies that treat insurance not as a procurement afterthought, but as part of AI governance, contracting, and commercialization discipline.

Ketaki Borkar is Senior Director, North America P&C, Sapiens.

Kim Harris is Business Analysis Manager, North America P&C, Sapiens.

Every Voice Matters: How AI-Powered Voice Translation is Improving Insurance Claims Operations

SYED RAZA, FSA, MAAA

The insurance claims process has long been recognized as one of the most critical touchpoints between insurers and their customers. It is the moment when policies transform from abstract promises into tangible support, when customer satisfaction is won or lost, and when operational efficiency directly impacts both experience and profitability. Yet despite decades of technological advancement, the claims process remains fraught with friction, particularly when language barriers enter the equation.

AI-powered voice translation represents a transformative technology poised to reshape how insurers handle claims across linguistic divides. By enabling real-time, accurate translation during voice interactions, this technology promises to accelerate processing times, enhance customer experience, reduce costs, and expand market reach. As these systems mature and adoption accelerates, they are positioned to become not merely incremental improvements but fundamental reimagining of how multilingual claims operations function.

The Language Barrier Challenge in Claims

Language barriers in insurance claims create cascading problems that affect every stakeholder. For customers filing claims in their non-native language, the process becomes exponentially more stressful. Accident victims or disaster survivors already navigating trauma must now struggle to communicate complex details about damages, injuries, and circumstances in an unfamiliar tongue. Misunderstandings multiply. Critical information goes uncommunicated. Frustration builds as simple exchanges require repeated clarification.

For insurers, language barriers translate directly into operational inefficiencies and elevated costs. Claims that should require a single phone call instead demand multiple interactions. Specialized multilingual staff become bottlenecks, limiting scalability and creating coverage gaps outside standard business hours. Translation services, whether human interpreters or document translation, add expense and delay. Quality suffers as nuances are lost and details become garbled through multiple intermediaries.

In an industry built on the principle of shared risk and mutual support, AI voice translation helps insurers live up to that promise across all communities they serve.

How AI Voice Translation Transforms Claims Processing

AI-powered voice translation leverages advances in machine learning, natural language processing, and speech recognition to enable real-time, bidirectional translation during voice conversations. Unlike traditional translation services that operate sequentially and often asynchronously, these systems translate speech nearly instantaneously, allowing natural conversational flow between parties speaking different languages.

The technology operates through several integrated components. Advanced speech recognition systems convert spoken language into text with increasing accuracy, even handling accents, dialects, and industry-specific terminology. Neural machine translation models translate the text into the target language with contextual understanding that captures meaning rather than merely substituting words. Finally, text-to-speech systems convert the translation back into natural-sounding speech, completing the loop.

A claims adjuster speaking only English can conduct a detailed first notice of loss interview with a Spanish-speaking claimant, with both parties speaking naturally and understanding each other in real-time. The technology captures not just words but context, enabling accurate communication of technical insurance concepts, damage descriptions, and procedural requirements.

Accelerating Claims Resolution

Speed is paramount in claims processing. Delayed settlements create customer dissatisfaction, increase costs through additional handling, and can escalate disputes. AI voice translation accelerates claims in several key ways:

- First, it eliminates scheduling delays associated with finding available interpreters or multilingual staff. Claims can proceed immediately upon customer contact rather than waiting hours or days for appropriate resources. This is particularly valuable for first notice of loss, where rapid response significantly improves customer experience and enables faster damage mitigation.
- Second, real-time translation enables single-call resolution for straightforward claims. Rather than requiring multiple interactions filtered through different intermediaries, the adjuster and claimant can complete the entire process in one conversation. Information gathering, explanation of coverage, discussion of next steps, and even settlement negotiation can occur without artificial breaks or delays.
- Third, the technology reduces miscommunication that commonly extends claims cycles. When messages pass through human interpreters or written translation, details are frequently lost, garbled, or misunderstood, necessitating follow-up clarification. Real-time AI translation, particularly systems that provide confidence scores and flag potential ambiguities, reduces these errors and enables more accurate communication on the first attempt.

For complex claims requiring extended investigation, AI voice translation enables ongoing direct communication between adjusters and claimants throughout the process. Status updates, requests for additional information, and settlement discussions occur fluidly without the friction of arranging translation services for each interaction. This continuous communication prevents the common problem of multilingual claims languishing due to communication difficulties.

Enhancing Customer Experience

AI voice translation fundamentally humanizes the multilingual claims experience. Customers can speak in their native language, expressing themselves naturally and completely rather than struggling with limited vocabulary or grammatical uncertainty. The cognitive load decreases dramatically, enabling them to focus on providing accurate information rather than on linguistic mechanics.

The technology also preserves dignity and reduces the vulnerability many customers feel when forced to communicate about sensitive or traumatic situations in a non-native language. Speaking one's mother tongue in moments of stress provides psychological comfort and enables more complete emotional expression. For trauma situations like serious accidents or catastrophic property losses, this psychological dimension significantly improves the overall experience.

Improving Accuracy and Compliance

Accuracy in claims communication is not merely about customer service—it carries legal, regulatory, and financial implications. Misunderstandings about coverage, settlement amounts, or procedural requirements can lead to disputes, complaints, or regulatory violations. Language barriers significantly increase these risks.

Modern neural translation models achieve remarkably high accuracy, particularly for common languages and standard insurance terminology. They understand context and can distinguish between homonyms or resolve ambiguous phrases based on surrounding conversation. Many systems specifically trained on insurance language and claims scenarios perform even better with industry-specific terminology.

Challenges and Considerations

Translation accuracy, while impressive, is not perfect. Certain languages, dialects, and accents present difficulties. Technical terminology, regional expressions, and cultural nuances can be misaddressed. Critical claims decisions should not rely solely on AI translation without human oversight.

Cultural competency extends beyond mere language translation. Effective claims handling requires understanding cultural norms around communication style, authority, family involvement, and expectations. AI translation provides linguistic access but does not automatically confer cultural sensitivity. Adjusters still need training in cultural competence to serve diverse populations effectively.

The Road Ahead

AI voice translation technology continues to advance rapidly. Accuracy improves with larger training datasets and more sophisticated models. New capabilities emerge regularly, emotion detection, accent adaptation, real-time dialect adjustment. Integration with other AI technologies creates even more powerful applications.

The convergence of voice translation with other AI claims tools promises further transformation. Imagine claims where AI voice translation enables the conversation, AI damage assessment analyzes photos in real-time, AI fraud detection monitors for red flags, and AI settlement recommendations optimize outcomes, all working seamlessly together to create rapid, accurate, customer-friendly claims resolution regardless of language.

In an industry built on the principle of shared risk and mutual support, AI voice translation helps insurers live up to that promise across all communities they serve, in every language those communities speak.

Syed Raza, FSA, MAAA is Founder & Principal, Actuarial 360 Solutions.



ACTUARIAL INTELLIGENCE BULLETIN

The Rise of the High Throughput Operator

MIKE WALSH

For most of modern knowledge work, the defining anxiety has been simple and persistent: am I doing enough? Enough hours, enough output, enough visible effort to justify my role and my compensation. Performance was measured in activity, and productivity was largely a function of how effectively human effort could be applied to a problem. But what changes when effort is no longer the constraint? When intelligence itself becomes elastic, abundant, and on demand, the question shifts. The rise of the token economy is often treated as a technical or financial detail, but it is something more revealing. It is emerging as a new measure of productivity, not in terms of effort, but in terms of leverage.

The early signals are striking. Some of the most sophisticated practitioners now worry less about cost discipline and more about underutilization. Andrej Karpathy has described feeling “nervous” when he does not fully exhaust his AI token allocation, treating unused capacity as lost opportunity rather than efficiency. Nvidia CEO Jensen Huang is even more explicit: “If that \$500,000 engineer did not consume at least \$250,000 worth of tokens, I am going to be

deeply alarmed.”¹⁰ Failing to deploy AI is no longer prudence. It is underperformance. The benchmark is shifting from how much work an individual completes to how much intelligence they can bring to bear.

This shift is best understood as a change in constraints. For decades, the bottleneck in knowledge work was human effort. Organizations were built to allocate tasks, coordinate people, and extract efficiency from limited time and attention. Generative AI introduces a different dynamic. Intelligence, once scarce and tightly coupled to individuals, becomes fluid and scalable. The limiting factor moves again. It is no longer what the system can do, but how effectively humans can direct it. On the No Priors podcast, Karpathy pointed out that the primary constraint in engineering work is no longer compute capacity. “It’s not about flops... it’s about tokens. What is your token throughput and what token throughput do you command?”¹¹ Performance is no longer defined by effort, but by the ability to direct large flows of machine intelligence toward meaningful outcomes. The implication is clear. If you cannot do this effectively, you become the constraint.

The High Throughput Operator is not the person who knows the most or works the hardest, but the one who can effectively coordinate the largest volume of intelligence.

In practice, this is already reshaping how work gets done. Tasks that once defined expertise are increasingly delegated to AI systems, while humans focus on structuring problems, distributing work across multiple agents, and integrating results. The role begins to resemble orchestration more than execution. Instead of writing code, drafting documents, or performing analysis step by step, individuals manage flows of machine-generated output across several tools at once, intervening at key moments to guide direction and ensure coherence. Less like a worker, and more like a system designer.

This is the emergence of a new archetype of performance: *the high throughput operator*. This is not the person who knows the most or works the hardest, but the one who can effectively coordinate the largest volume of intelligence. Their advantage lies in how they frame problems, how they allocate tasks between human and machine, and how they maintain quality across an expanding surface area of output. They treat AI not as a tool to be occasionally consulted, but as an always on cognitive infrastructure. Their contribution is not measured in tasks completed, but in systems directed.

In this environment, expertise does not disappear, but it changes shape. Knowledge becomes a multiplier rather than a primary source of value. The critical skill is judgment, knowing how to break problems into machine executable components, how to design workflows that produce useful results, and how to evaluate those results before errors compound. This is where cognitive leverage becomes the defining concept. Cognitive leverage is the ability to generate disproportionate value from a relatively small amount of human input. It is the difference between doing more and making more happen. A highly leveraged individual can take a complex objective, distribute the work across a network of AI systems, and recombine the outputs into something coherent and valuable. Tokens enable this process, but they do not determine its effectiveness. That depends on how well the system is designed and governed.

This introduces a familiar tension. Tokens are both a cost and a capability. The instinct to minimize usage is understandable, but it risks constraining the very resource that drives productivity. History suggests that

¹⁰ Kif Leswing, “Jensen Huang Says 500,000 Engineers Can Use 250,000 AI Tokens Thanks to Nvidia Compute,” *Business Insider*, March 2026, <https://www.businessinsider.com/jensen-huang-500k-engineers-250k-ai-tokens-nvidia-compute-2026-3>.

¹¹ “No Priors: Artificial Intelligence, Technology, Startups,” *Apple Podcasts*, <https://podcasts.apple.com/no/podcast/no-priors-artificial-intelligence-technology-startups/id1668002688>.

organizations that expand into new forms of abundance outperform those that optimize too early. Electrification created advantage not because power was cheap, but because it enabled entirely new ways of organizing production. Cloud computing followed the same pattern. It won not on cost efficiency alone, but on the ability to experiment and scale. The same logic now applies to intelligence. The question is not how much is consumed, but how effectively it is deployed.

At the same time, the labor market is beginning to adjust. The routine, structured tasks that once defined entry-level roles are among the first to be automated, reducing demand for junior positions while increasing the premium on those who can operate at a higher level of abstraction. This creates a subtle but important shift. The pathway to expertise, historically built on repetition and incremental skill acquisition, is narrowing just as the need for high-quality judgment expands. Without a deliberate approach to talent development, organizations may find themselves with more intelligence than they can direct, but fewer people capable of directing it.

As models improve and costs decline, the constraint will move again. Access to tokens will matter less. The scarce resource will be judgment, the ability to ask better questions, structure problems, and intervene at the right moments. In that world, performance is no longer about what you produce, but what you can direct. Leverage becomes the defining metric.

The implication is stark. When intelligence is abundant, underutilization becomes the new form of inefficiency.¹² Not using what is available is no longer a sign of discipline, but of misalignment. The organizations that struggle will not be those that lack access to AI, but those that fail to reorganize around it. And the individuals who fall behind will not be those who lack effort, but those who fail to expand their capacity to direct it.

Mike Walsh is Global Futurist at Mike-walsh.com



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 May 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.

¹² Michael J. Walsh and Nitin Mittal, *Abundant Intelligence: How Digital Labor Will Rewrite the Rules of Business*, Harvard Business Review Press, January 12, 2027.

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