

ACTUARIAL INTELLIGENCE BULLETIN



November 2025

From Hype to Help: Practical AI for Actuarial Workflows	2
The Human Side of AI: Ethics, Risk, and Responsibility	4
Turning Up the Heat: Understanding Temperature in Large Language Models	7
Upcoming Webcasts.....	9
When AI Becomes a Colleague: Agentic Intelligence in the Actuarial World.....	9
Management of AI Risks.....	11
The LM Arena: A Crowdsourced AI Benchmark.....	13
New AI Subgroup Launch!	16
Is This the Last Invention or the Next Valuation?	17
How to Insure AI? AI Insurance is Already Here	20
Artificial Intelligence Research	22
Vibe Coding – Gutenberg’s Printing Press or Pandora’s Box.....	22
Creating and Using Synthetic Data	24
The Fifth Kind of Actuary: Expanding Our Thinking for a VUCA World	27
Editorial Committee	29
About the Society of Actuaries Research Institute	30

Welcome to the November 2025 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.

From Hype to Help: Practical AI for Actuarial Workflows

MAHIMA KHANDELWAL

Picture this: you're in the middle of quarter-end chaos—emails pinging, numbers shifting, and your report deadline looming. You press the big imaginary “Make my report” button: the tool asks GenAI a few targeted questions about the numbers, writes a board-ready paragraph, and drops it (charts and all) straight into your report.

That's the power of GenAI. Think of it as a witty, caffeinated intern—except this intern never sleeps, never forgets, and is powered by APIs and automation that handle the grunt work in the background.

For actuaries, whose world is full of models, documentation, and regulatory reporting cycles, this isn't a distant dream—it's already starting to happen. But as exciting as it sounds, one truth runs through everything: **speed without trust is meaningless**. Governance isn't an afterthought—it's the foundation that allows these tools to scale responsibly.

So, the real question isn't “Will AI impact actuarial work?” but “How quickly can we adapt—and how do we build trust as we do so?”

Automation unlocks scale. Instead of 10 tests, an actuary can run 200, surface the few that matter, and iterate with treasury the same day. Errors shrink, reviews speed up, and actuaries focus on judgment, not formatting. Guardrails like prompt templates and plausibility checks keep outputs consistent and safe. In short: automation doesn't replace judgment—it amplifies it at scale.

Speaking Plainly: LLMs, APIs, and Why They Matter

At its core, a Large Language Model (LLM) is just a super-powered text predictor. It takes in words (your prompt) and predicts the next ones with uncanny fluency. But the magic emerges when paired with APIs—the pipes that let it talk to your existing systems.

Here's the simple picture:

- **You:** Ask a question (“Summarize reserve movement for the quarter”).
- **API (the connector):** Pulls the right data, adds context (documents, charts, retrieved snippets), enforces security and logging, and calls the LLM.
- **LLM:** Writes the summary, tailored to your audience.

That's the core pattern. It's not about replacing actuarial judgment—it's about giving actuaries a **power tool** that connects what you already have to what you need, faster.

A Hybrid Actuarial Workflow

Anita, a valuation actuary, types a simple instruction:

“Generate 200 plausible 20-year scenarios with a flu-like mortality wave and a two-year lapse spike after an employer closure.”

Behind the scenes, the system calls the LLM via an API and returns neat scenario headlines, a one-line rationale, and numeric parameters the valuation engine can plug in. What Anita sees looks like the first draft from a sharp analyst—done in minutes.

Her first run taught Anita a blunt lesson: how you ask matters. “Make extreme scenarios” produced vague, hard-to-use stories. When she changed the brief to: *“Return a title, a one-line rationale, and three plug-in numbers—mortality multiplier, lapse multiplier, and a one-line interest path”* the outputs became precise and actionable.

Prompting isn’t a trick—it’s the difference between a usable draft and noisy output you must clean up. Better prompts give you consistency, speed up review, reduce hallucinations, and make results reproducible; poor prompts deliver—quite literally—garbage in, garbage out.

Quick tips Anita follows: specify the exact format, show a tiny example, list required fields, and version your prompt templates so they’re auditable.

When Anita asks for a scenario that reflects an employer-closure lapse spike, the system doesn’t guess corporate realities—it looks them up. Behind the scenes, relevant company documents (policy wording, reinsurance treaties, internal mortality studies) are converted into embeddings and stored in a vector database. The RAG pattern then retrieves the most relevant snippets and supplies them to the LLM as context, so each scenario can reference the exact clause or study that motivated it.

For Anita, that means the narratives aren’t generic: they cite firm evidence, reduce hallucination risk, and make the scenarios easier to defend to auditors and management.

Anita knows from experience that one big request “Generate scenarios, give me the numbers, and suggest hedges” often produces a jumble: some parts useful, others unusable. Instead, she breaks it down. First, the model creates clear scenario stories. Next, it extracts the exact multipliers her valuation engine needs. Finally, it proposes a shortlist of hedge ideas with plain-English rationales. By chaining the steps, Anita turns one messy black box into a series of clean, reviewable outputs. Each step can be checked and corrected before moving on—making the process faster, more accurate, and far more trustworthy.

Finally, Anita’s team samples outputs, checks for coherence (no negative interest rates, no unrealistic mortality spikes), and signs off. AI drafts, compares, and summarizes like a pro—but Anita still applies her actuarial expertise to check assumptions and ensure the model’s integrity. That human-AI partnership is the sweet spot.

It sounds complex, but in practice it feels like magic—because all Anita sees is a button that drafts her first scenarios for her.

Why Automate?

AI can crunch volumes of data in seconds—but it won’t catch a regulatory nuance or interpret a mortality trend without Anita’s judgment. That’s why she automates: not to replace expertise, but to make answers usable, repeatable, and defensible. Her pipeline builds a searchable scenario catalogue—stories, plug-in numbers, and the supporting snippets—so results can be re-run, compared, and audited with a clear provenance trail.

Automation also unlocks scale. Instead of 10 tests, Anita can run 200, surface the few that matter, and iterate with treasury the same day. Errors shrink, reviews speed up, and actuaries focus on judgment, not formatting. Guardrails like prompt templates and plausibility checks keep outputs consistent and safe. In short: automation doesn’t replace judgment—it amplifies it at scale.

AI’s Real Value for Actuaries

GenAI’s real value isn’t in replacing actuaries but amplifying them—turning data into clarity and freeing time for judgment and insight. The next step? Scaling responsibly with trust, governance, and transparency. Stay tuned for Part 2 on building AI workflows that actuaries can truly trust.

A version of this article has recently appeared in the third 2025 edition of the Society of Actuaries Knowledge News Quarterly, India Edition.

Mahima Khandelwal is Senior Consultant at EY in Nagpur, Maharashtra, India



ACTUARIAL INTELLIGENCE BULLETIN

The Human Side of AI: Ethics, Risk, and Responsibility

CARLOS AROCHA, FSA

AI has become one of the most discussed topics of our time. From voice recognition on smartphones to recommendation engines, and from medical diagnostic tools to fraud detection, AI shapes daily life in both visible and invisible ways. Yet despite its presence, many people still misunderstand what AI is, how it works, and why it matters for society.

I recently explored these questions in a live discussion on *Dialogues with My Psychoanalyst*,¹ a weekly program on Radio El Heraldo in Mexico City. As an actuary, I shared how AI affects a profession centered on risk and uncertainty, and why it must be seen not only as a technological advance but as a decision-making tool with profound human consequences.

This article reflects on three themes from that conversation: what AI truly means, how it is shaping daily life, and why it raises pressing ethical questions.

Defining AI: Beyond the Myths

Popular culture often portrays AI as human-like robots. The reality is simpler but no less powerful: AI systems use algorithms and data to perform tasks requiring human intelligence, such as recognizing patterns, predicting outcomes, and adapting to new information.

Most applications today are “narrow AI,” built for specialized tasks like image recognition, translation, or traffic prediction. “General AI,” capable of reasoning across domains, remains far from reality. A central feature of AI is machine learning (ML), where systems learn from data rather than explicit instructions. For instance, a program can distinguish cats from dogs not because it was told how, but because it trained on thousands of images.

For actuaries, ML represents continuity as well as change. We have long relied on statistics, probability, and large datasets to analyse risk. What is new is the scale and speed at which AI can detect patterns and improve accuracy.

¹ A podcast version of the episode (in Spanish) can be found at <https://open.spotify.com/episode/7wrMteQN1T2DJgqDzOwoq5?si=7dc71a2bb5574954>

AI in Daily Life: Opportunities and Risks

AI is embedded in everyday experiences, often unnoticed.

- Healthcare: AI analyses scans, predicts patient deterioration, and supports personalized treatment.
- Finance and insurance: Banks detect fraud using AI, while insurers refine risk assessments. Actuaries use ML to uncover patterns in claims and policyholder behaviour.
- Transportation: Navigation apps optimize travel, while autonomous vehicles promise to reduce human error.
- Retail and media: Recommendation engines shape consumer choices and cultural consumption.
- Public services: Governments employ algorithms for traffic management and resource allocation.

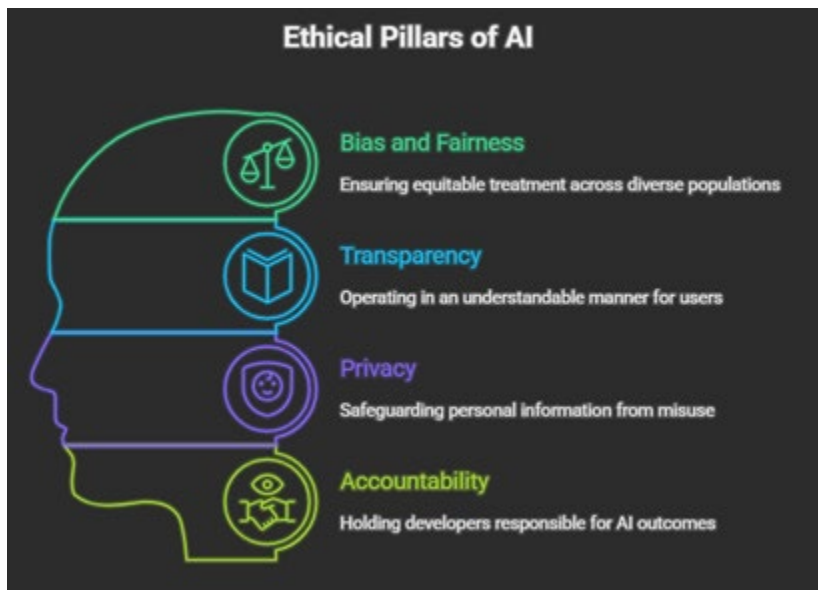
These applications show that AI is not reserved for specialists. Even without technical knowledge, individuals experience AI-driven decisions—from which job postings appear online to how insurance premiums are set.

Ethics and Responsibility in AI

The rise of AI also raises questions about fairness, accountability, and trust.

- Bias and discrimination: AI reflects the data it learns from, which often includes historical inequalities. A hiring tool, for example, may unintentionally favor candidates resembling past employees.
- Opacity: Many AI systems, particularly deep learning models, operate as “black boxes.” When used in insurance lack of transparency undermines trust.
- Privacy: AI thrives on data, but much of it is personal. Balancing personalization with protection of individual rights is a critical challenge.
- Accountability: If an AI system makes an error—such as a self-driving car causing an accident—who is responsible: the developer, the user, or the machine?
- These issues call for governance frameworks comparable to those actuaries apply in solvency and capital adequacy. Transparency, human oversight, and ethical guidelines are essential to ensure AI systems serve society responsibly.

Data carries history, and history carries bias.



Source: Arocha & Associates GmbH

The Actuarial Lens: Risk and Uncertainty

For actuaries, AI strengthens traditional practices demanding vigilance. Examples include:

- Climate risk: Traditional models rely on historical data, while AI can integrate satellite imagery and sensor data to provide more detailed assessments. This supports insurers, governments, and communities in preparing for extreme events.
- Healthcare insurance: ML models identify early signs of chronic disease, guiding both underwriting and preventive care.

Yet AI is no panacea. Models are only as reliable as the data and assumptions behind them. Blind reliance on algorithms introduces systemic risks. The actuarial profession's emphasis on ethics, professionalism, and standards offers safeguards against these dangers.

A Shared Future

AI will continue to advance, but its trajectory is not predetermined. The direction it takes depends on the choices made by developers, policymakers, businesses, and the public. The challenge is to maximize the benefits of AI while managing its risks. Education will play a vital role. For the general public, a basic understanding of AI helps demystify it and encourages informed debate. For professionals, cross-disciplinary knowledge that combines technical skills with ethical, social, and economic perspectives will be essential.

Artificial intelligence is a complement to human judgment, not a replacement.

As an actuary, I see AI as an opportunity to enhance our decision-making under uncertainty. It is a powerful complement to human judgment, not a replacement. The most responsible path forward is to recognize both its potential and its limitations.

Conclusion

Artificial intelligence is no longer an abstract concept confined to research labs. It is a practical tool shaping how we live, work, and interact. It promises improved healthcare, safer transport, and more efficient use of resources. But it also compels society to confront difficult questions of fairness, privacy, and responsibility.

By examining AI through the lens of ethics, risk, and responsibility, we see its dual nature: both opportunity and challenge. Whether we are actuaries, doctors, teachers, or consumers, we all have a stake in shaping AI's future. Ultimately, the conversation about AI is a conversation about us—our values, our choices, and our vision for society.

I thank the psychoanalysts of *Dialogues with My Psychoanalyst* for the opportunity to share these reflections with their thoughtful audience.

Carlos Arocha, FSA, is Founder and Managing Partner of Arocha & Associates GmbH. He can be reached at ca@ArochaAndAssociates.ch

**ACTUARIAL
INTELLIGENCE
BULLETIN**

Turning Up the Heat: Understanding Temperature in Large Language Models

YUKKI YEUNG, FSA MAAA

When you ask a question to a system like ChatGPT, Gemini, or Claude, the words you see aren't just popping out at random. They're drawn from probabilities calculated by billions of parameters, influenced by a simple but surprisingly powerful inference setting, one of them called **temperature**.

What Exactly Is Temperature?

Think of temperature as the model's "creativity knob." Technically, it adjusts how sharply or loosely the model treats its probability distribution when picking the next word. With a low temperature, the model sticks to the most likely option and feels more certain with high probabilities. Push the temperature higher, and the model starts giving fewer probable words a chance, which can make its answers sound more adventurous, but also increase chance of hallucination.

In practice:

- **Low temperature (0.1 to 0.3):** Output is consistent, concise, and almost certain in tone.
- **High temperature (0.7 to 1.0):** Responses feel freer, more playful, sometimes offbeat.
- **Middle ground (around 0.5):** Often the default is a mix of steadiness and variety.

Some people say higher temperature "just adds noise." That's not quite right. It's closer to a tradeoff between high belief and exploration. At lower settings, you'll usually get one "probable" answer, which is handy for things like coding or financial calculations. At higher settings, the model experiments. That's when it's more likely to suggest an unusual analogy for a risk report, or to spin out a quirky poem.

The temperature dial isn't magical. Too low, and the text can start sounding like the same sentence on repeat. Too high, and you may get answers that drift or don't quite make sense.

Of course, the dial isn't magical. Too low, and the text can start sounding like the same sentence on repeat. Too high, and you may get answers that drift or don't quite make sense.

Everyday Use Cases

Here's how people use it in practice:

- **Low temperature:** Customer service scripts, legal write-ups, actuarial calculations, technical documentation. Anywhere accuracy trumps flair.
- **High temperature:** Marketing copy, naming brainstorming, storytelling, or scenario generation. Basically, when you want options instead of "the one right answer."
- **Middle range:** Good default for emails, casual summaries, or mixed audiences.

How Different LLMs Handle It

The setting works roughly the same across models, though each has its own flavor.

- **ChatGPT:** Defaults conservative, but API ranges from 0 to 2.
- **Claude:** Even at higher temps, it tends to sound cautious and measured, which some people appreciate in sensitive fields.
- **Gemini:** Temperature influences not just text but multimodal responses (e.g., mixing text with images or code snippets).

Why Actuaries Should Care

This isn't just a quirk for techies. For actuaries, temperature can shape real workflows and reproducibility:

1. **Model documentation:** Use low temps to produce audit-ready write-ups of valuation methods or pricing assumptions.
2. **Training and compliance:** Draft consistent answers for regulatory Q&A at low temps but raise it for role-play scenarios in professionalism training.
3. **Product development:** Low temps for precise competitor summaries, high temps for testing creative new benefit ideas.
4. **ERM or audit risk testing:** Low temps for checklist generation, higher ones for imagining extreme scenarios.
5. **Stakeholder communication:** Low temps for clear explanations of results, high temps for coming up with analogies that resonate with executives.

Do Regular Users Need to Worry About This?

If you're not fiddling with an API, you still bump into temperature effects without realizing it. Ever notice how sometimes the model gives you a very direct answer, and other times it riffs? That's temperature showing through.

A few tips:

- **Manage expectations:** Don't be surprised if the same question gives slightly different answers.
- **Prompt style matters:** Asking "Give me three creative options" nudges the model toward higher-temperature behavior.
- **Verify:** Low-temperature outputs may *sound* reliable, but they're not immune to mistakes. High-temperature ones might spark ideas but need reality checks.

Final Thought

Temperature may look like just a single slider, but it can change how a model feels, monotone robot on one end, poet on the other. For most professionals, the trick isn't knowing the math; it's knowing how to guide the model into the "just right" zone for the task at hand.

YUKKI YEUNG, FSA MAAA the functional lead for actuarial, AI, and data analytics audits across North America at Legal and General

Ed. Note: There are more dials like Top-P (Nucleus Sampling), Top-K Sampling, Seed, Frequency Penalty (various penalties) that we will address in future articles.

Upcoming Webcasts

Bridging the Gap: Actuaries and Data Scientists Work Together Webcast

November 14, 2025, 12:00–1:30 PM ET

As data science continues to expand across industries, actuaries are increasingly called upon to collaborate with data scientists to unlock new insights and enhance decision-making. This webcast explores how actuaries can effectively partner with data science teams by building a shared understanding of key terminology, aligning on project goals, and leveraging complementary skill sets.

<https://www.soa.org/prof-dev/webcasts/2025-btg-act-and-data-scientists/>

Beyond the Prompt: Custom AI Assistants Webcast

November 21, 2025, 3:00–5:00 PM ET

This session introduces you to the next level of gen AI: AI Assistants (also known as Custom GPTs, Gems, or Bots). An AI Assistant is a specialized, tailored version of a Large Language Model (LLM) that allows the user to produce a persistent, pre-programmed tool designed to handle complex, repetitive professional tasks with guaranteed consistency, accuracy, and alignment.

<https://www.soa.org/prof-dev/webcasts/2025-beyond-prompt-custom-ai/>

ACTUARIAL INTELLIGENCE BULLETIN

When AI Becomes a Colleague: Agentic Intelligence in the Actuarial World

SALMA GALVAN SOLANO

Artificial intelligence (AI) is now part of many processes, from chatbots to code generation, and although AI is not actually a new concept—it has existed since the 1950s in different shapes such as machine learning, deep learning or neural networks—it gained a lot of relevance with the launch of Generative AI systems. Now, a new concept has started to solidify: Agentic AI. This new generation of AI represents a paradigm shift: systems not only capable of performing tasks that amaze us, but also of doing so autonomously, guided only by a predefined goal.

Agentic AI introduces a dynamic approach where multiple agents collaborate, share information, and learn from experience to achieve complex objectives.

An AI agent can be understood as an intelligent system that perceives, reasons, and acts within a defined environment to accomplish specific objectives. I like to imagine an AI agent as an intelligent workflow that doesn't need to be programmed step by step. You only need to define the goal and provide the tools or resources available, and the agent will interpret this information, plan a course of action, and execute decisions to move toward that goal. The main difference against traditional AI models is their ability to plan, make decisions, and execute tasks autonomously. The concept of Agentic AI takes these capabilities even further by introducing a dynamic approach where multiple agents collaborate, share information, and learn

from experience to achieve complex objectives, marking a new stage in the evolution of artificial intelligence.

Although these agents appear intelligent and capable of reason, there are several technological components that make this illusion possible in practice:

- **Prompt Engineering:** The process of designing prompts to obtain the desired behavior from an AI agent. This defines how the goal is communicated to the agent and how it interprets the instructions.
- **Fine Tuning:** A technique that allows a model to learn specific behaviors. It starts with a pre-trained model and uses a fine-tuning dataset containing example queries paired with the desired responses. The fine-tuning process adjusts the model's internal parameters, so it becomes more likely to generate outputs aligned with those examples.
- **Retrieval Augmented Generation (RAG):** A process that enables the agent to access external data sources and incorporate relevant information into its reasoning and responses.
- **Tool Use or Function Calling:** Allows the agent to execute real actions and connect to external APIs, transforming the agent from a passive expert into an active assistant.
- **Planning Frameworks:** Organize the sequence of steps the agent follows. There are two main approaches: flow engineering, which involves breaking down a task into a sequence of sub-tasks, and multi-agent systems, which take a modular approach by delegating different roles to specialized sub-agents.

When these techniques come together, they give rise to what is known as agentic behavior, empowering AI systems to perform sophisticated tasks independently, collaborate effectively, solve problems, and execute actions with little human guidance.

In the actuarial field, the potential of Agentic AI can already be illustrated through practical use cases. For instance, an AI agent can automatically generate the documentation of an actuarial model by analyzing its structure, assumptions and parameters producing a comprehensive report with such a detailed level that you could even reproduce the code from the documentation alone. Another example is an agent designed to translate an Excel-based model into a more robust modeling environment by interpreting formulas, identifying data dependencies, and creating a hierarchy of the data. The agent can recreate the model into a new platform by generating code depending on the software that replicates the behavior from the excel-based model. These applications demonstrate how Agentic AI can enhance efficiency, reduce operational risk and free up actuaries to focus on higher-value analytical tasks.

As Agentic AI continues to evolve, its integration into the actuarial field will likely transform the way analytical and operational tasks are performed. Rather than replacing actuarial expertise, these intelligent systems will enhance it, and it is crucial that we, as actuaries, learn how to work alongside AI agents to operate in ways that are both more efficient and more strategic. Agentic AI is not just a technological innovation; it is the evolution from static tools to dynamic collaborators.

Salma Galvan Solano is a Consultant for WTW in Mexico City

Management of AI Risks

AN INTERVIEW WITH THE JOINT RISK MANAGEMENT SECTION COUNCIL

AIB: What do you think of when we talk about AI risk? What are your concerns?

Alexander Tall: Two sides of the coin. One is the actual risk of using AI in a less controlled way—using it inappropriately, hallucinations creeping into reporting or documentation.

The other is the opposite: the risk of being left behind. Not being able to compete because you're not using AI the way competitors are.

Joan Barrett: Inability to validate findings. I'm not sure about the input, the process, or the output—and how to validate that over time. The black box nature of AI is concerning.

Monojit Samanta: There's also the external AI threat. For example, phishing attempts, KYC challenges. AI can be used by bad actors to disrupt processes. Hackers are more efficient now—AI can generate bad code without needing to learn to code.

Bryan Liu: The unknown impact of AI. The known effects of AI on risk management, or the ways it can pose a challenge, are important. However, as the technology develops, there are risks that have not yet emerged. For example, if AI suddenly creates a huge financial crisis, that's a risk we're not currently capturing.

Alexander Tall: Another concern is reduction of the role of actuaries as things get outsourced to AI.

Emily Li: And with AI tools, junior staff are competing with AI instead of being trained. Managers may not give junior staff opportunities to learn and grow. AI gets smarter, but juniors aren't developing.

Monojit Samanta: That affects succession planning. If you don't do the grunt work, how do you learn the business, the rules of thumb? If analysts jump straight to associate or manager level, without grounding, it's a problem. Unless the whole company is run by AI, which I don't see happening soon.

Aly Moosa: We've talked about whether AI will be able to explain things well enough. But I personally fear losing out on the associated human oversight. We need to be able to explain the results like humans do to ensure there is an understanding of the underlying aspects.

We are using AI so often for decision making, and when we talk about high-impact outcomes, our dependency may be too much. We need to ensure we're being conscious and prudent in what we're taking in and how we're separating out the biases associated to it.

Bryan Liu: There is also the strategic risk of doing too much with AI before the technology is at a stage where it can have a bigger, better impact. Spending too much time and effort trying to fit AI into your organization prematurely is a risk.

AIB: Do these create new risk categories, or do they fit existing structures?

Emily Li: Most fit into existing taxonomy. Succession planning falls under human capital risk. Cybersecurity risks cover bad actors. Model risk covers interpretability and bias. It's more about expanding existing categories to account for AI.

AIB: How do you distinguish real AI risks from hype?

Monojit Samanta: It comes down to exposure. If AI is integrated into your systems, risk depends on how critical that model is. Strategy risk depends on your industry—if peers are advancing quickly with AI, the risk of falling behind is real. Operational risk is where media hype plays in—when you hear of hacks, it’s not always clear if AI was involved. You need specialists to assess how real the risk is for your company.

Emily Li: Companies view AI differently. If you’re still in exploration, it may be an emerging risk. If you only use AI for drafting emails, the exposure is low. But if you use AI for experience studies, that’s higher exposure. Companies weigh likelihood, impact, and velocity of adoption.

Aly Moosa: When discussing risk, we sometimes think of the worst-case scenario and post things about AI in an extreme or catastrophic way. But this is such a new risk that, over time, we’ll be able to get a more measurable understanding of it. Right now, the hype is so uncertain, and we think things can be so much worse, but I don’t know if that’s true.

AIB: Do you have risk metrics that are sensitive to AI use?

Emily Li: Mainly reflected in IT metrics and model risk metrics.

Bryan Liu: Anything on the operational side of things, like any existing operational risk (op risk) metrics, should capture AI use. For example, model risk, fraud risk, or human capital risk metrics can be used.

Aly Moosa: Given how AI hallucinates, there could be a measure of the proportion of AI outputs that are factually incorrect or nonsensical. That could be the metric where people say, “This tool is 80% believable,” or “This tool hallucinates 15% of the time.”

AIB: What are companies doing to mitigate AI risks?

Aly Moosa: We have established governance on the use of tools like Google Gemini and ChatGPT. This includes policies on what you can upload, what you can ask, and what data you can share.

Companies are also developing an ethics framework around the principles of fair, transparent, and accountable use of AI. Eventually, companies will likely have an AI governance committee.

Bryan Liu: Companies are trying to mitigate AI risk by exploring AI and spending more time on it, so they don’t lag their competitors. They are trying to not be left behind.

Emily Li: It depends. For succession planning, mitigation is about training. For quality of AI tools, it’s testing, interpretability analysis, validation. No standardized approach yet—companies are still exploring.

On training, some companies use a hybrid approach: juniors work with AI, drawing conclusions together rather than relying only on AI.

Alexander Tall: Some companies prevent access to ChatGPT and other public LLMs from work computers to avoid uploading confidential information. Others use internal or secured third-party solutions.

Monojit Samanta: It takes expertise to assess whether AI is right. If you’re not familiar with a topic, it’s easy to believe something that’s inaccurate.

Joan Barrett FSA, MAAA is Partner, Axene Health Partners, LLC Tolland, Connecticut
Emily Li, FSA, MAAA, CERA, FRM Manager at Deloitte Consulting, LLP
Bryan Liu, FSA, FCIA Actuary at Milliman, Toronto
Aly Moosa, ASA, MAAA Head of Model Risk at National Life Group, Dallas, Texas
Monojit Samanta, FSA, FCIA Senior Actuarial Manager at TD Insurance, Toronto
Alexander Tall, FSA, MAAA Senior Manager Oliver Wyman, New York

ACTUARIAL INTELLIGENCE BULLETIN

The LM Arena: A Crowdsourced AI Benchmark

YU CHENG ASA AND ANDREW DILWORTH FSA, MAAA

When working with large language models (LLMs), model performance is clearly a critical consideration, yet the evaluation landscape can be opaque and difficult to navigate. A common approach is to test models using difficult multiple-choice questions, such as Massive Multitask Language Understanding (MMLU) or Graduate-Level Google-Proof Q&A (GPQA). However, it might not be clear to all users which benchmark is most applicable for their use case, and performance on any one specific metric may not necessarily reflect real-world utility.

The LM Arena leaderboard is not static—it evolves continuously based on user activity. As more head-to-head comparisons are submitted and new models are introduced, rankings shift accordingly.

The LM Arena offers a simple alternative that addresses this challenge. It is a publicly available leaderboard that ranks various AI models by **crowdsourcing user preferences**. Instead of relying on abstract metrics, it focuses on practical performance: how models respond to actual user prompts. This makes it a valuable resource for users who want to understand how models perform in everyday scenarios, without needing deep technical expertise and familiarity with numerous model evaluation techniques.

Note: Other important factors for model selection (such as cost, latency, or context length) are not explicitly considered in these rankings.

How the Ranking System Works

In the LM Arena, users submit a prompt, receive responses from two different models, and vote for the one they prefer. Model identities are hidden until after voting, ensuring that judgments are based solely on response quality. User votes are then translated into rankings by using the Elo rating system, originally developed to rank chess players. This system is applicable to any zero-sum competition and has been adapted for esports competitors, professional sports leagues, and now, AI model evaluation.

In the Elo framework, each AI model enters a head-to-head matchup with an existing rating. After a vote, the “winner” gains points from the “loser,” with the number of points transferred depending on their respective ratings. If a higher-rated model wins, fewer points are exchanged, as the outcome aligns with expectations. However, if a lower-rated model wins, more points are transferred, reflecting the “upset.” This dynamic system allows ratings to self-correct over time as more comparisons are made.

LM Arena's Impact

By exposing models to public evaluation, the LM Arena introduces a level of transparency and accountability rarely seen in AI development. Performance on the leaderboard becomes a form of reputational currency, encouraging developers to prioritize real-world effectiveness.

The LM Arena introduces a level of transparency and accountability that is rarely seen in AI development.

This competitive atmosphere helps drive innovation. Models are often fine-tuned or updated in response to Arena outcomes, with creators aiming to improve results across both general and specialized tasks. In this way, the leaderboard doesn't just reflect progress, it helps stimulate it.

The "Leaderboard Illusion"

Despite its usefulness, the LM Arena has drawn criticism from some in the AI community. A common concern is the so-called "leaderboard illusion," where a model's rank may not fully reflect its general capabilities. Strong performance in over-represented prompt types or specific linguistic contexts can lead to inflated scores. An additional concern could be demographic bias among voters. These issues are not unique to the LM Arena but are especially relevant in a system that relies heavily on public participation. Users are encouraged to interpret results with nuance and consider them as one part of a broader evaluation process.

The LM Arena's Commitment to Transparency

The LM Arena has taken concrete steps to ensure fairness and openness. Model identities are hidden during voting, pairings are randomized, and version tags are used to reflect model updates. This minimizes bias and allows for more meaningful comparisons over time. Detailed voting data is also made public, allowing researchers and users to analyze performance trends. These ongoing efforts help reinforce the platform's credibility and responsiveness to community feedback.

Current Rankings (as of September 16, 2025)

At the time of this writing (September 16, 2025), Google's *gemini-2.5-pro* and Anthropic's *claude-opus-4-1-20250805-thinking-16k* are tied for the top spot on the overall LM Arena leaderboard. Since rankings are updated dynamically, results may vary depending on when the leaderboard is viewed (a useful feature in a rapidly evolving field).

For users interested in general-purpose performance (ignoring cost and other constraints), these top-ranked models offer a strong starting point.

Specialized Rankings

Sometimes, users need models optimized for specific tasks rather than general use. The LM Arena addresses this need through specialized leaderboards in areas like web development, computer vision, and more. Below are the current leaders in each category (subject to change as new data is submitted), as well as brief descriptions of the distinct categories.

Category	Description	Top Two Models
Text	Model versatility, linguistic precision, and cultural context across text	Google's <i>gemini-2.5-pro</i> Anthropic's <i>claude-opus-4-1-20250805-thinking-16k</i>
Web Development ("WebDev")	Models specialized in web development tasks (HTML, CSS, and JavaScript)	OpenAI's <i>gpt-5-high</i> Anthropic's <i>claude-opus-4-1-20250805-thinking-16k</i>
Vision	Models that understand and process visual inputs	Google's <i>gemini-2.5-pro</i> OpenAI's <i>chatgpt-4o-latest-20250326</i>
Text-to-Image	Models that generate images based on prompt(s)	Bytedance's <i>seedream-4-high-res</i> Google's <i>gemini-2.5-flash-image-preview (nano-banana)</i>
Image Edit	Models that generate and edit images	Google's <i>gemini-2.5-flash-image-preview (nano-banana)</i> Bytedance's <i>seedream-4-high-res</i>
Search	Models with web search for real-time information, external knowledge, and grounded citations	OpenAI's <i>o3-search</i> OpenAI's <i>gpt-5-search</i>
Text-to-Video	Models that generate videos based on prompt(s)	Google's <i>veo3-fast-audio</i> Google's <i>veo3-audio</i>
Image-to-Video	Models that generate videos based on image(s)	Google's <i>veo3-audio</i> Google's <i>veo3-fast-audio</i>
Copilot ²	AI coding assistants	Anthropic's <i>Claude 3.5 Sonnet (06/20)</i> Deepseek AI's <i>Deepseek V2.5 (FIM)</i>

These results highlight the versatility of top models like Gemini, while also showcasing strong alternatives in specialized domains.

Dynamic Ranking Disclaimer

The LM Arena leaderboard is not static—it evolves continuously based on user activity. As more head-to-head comparisons are submitted and new models are introduced, rankings shift accordingly. This ensures that the results reflect the latest performance rather than a snapshot froze in time.

This dynamic structure is both a strength and a potential drawback. It captures real-time improvements and regressions in model behavior but also means that rankings can vary significantly from week to week. Users should be mindful that what holds true today may change quickly as new data flows in.

Conclusion: Human Judgment at the Core

The LM Arena reflects a fundamental shift in how AI performance is measured, moving from purely technical scoring to evaluation rooted in live interaction. Traditional benchmarks can offer valuable insights into specific capabilities, but they may fall short in capturing other qualities that matter in real-world use, such as helpfulness, nuance, and tone.

By placing human judgment at the center, the LM Arena offers a complementary perspective grounded in actual user experience. Crowdsourced evaluation is not perfect, but it provides a unique lens on progress: one that emphasizes not only what models can do in theory, but how they perform in the hands of real users.

² As of this writing, all arenas have been updated in the past few weeks except Copilot arena (last updated May 28, 2025).

As AI becomes more integrated into daily and professional life (including fields like actuarial science), the capacity to capture human-centered value will matter as much as traditional measures of accuracy or efficiency. The LM Arena serves as both a case study in innovative measurement and a glimpse into a future where model quality is defined not just by metrics, but by meaningful interaction.

*Yu Cheng ASA is Actuarial Advisor / Conseiller actuariel at Office of the chief Actuary / Bureau de l'actuaire en chef in Ottawa
Andrew Dilworth FSA, MAAA is Senior Director, Actuarial Services at UnitedHealthcare in Appleton, Wisconsin*

ACTUARIAL INTELLIGENCE BULLETIN

New AI Subgroup Launch!

ANDREW DILWORTH FSA, MAAA

The SOA Emerging Topics Community³ is pleased to announce the recent introduction of our **Artificial Intelligence subgroup**. The subgroup's purpose aligns with the third key priority of the SOA's strategic plan (**Leverage AI**), offering an opportunity for actuaries to develop AI expertise and promote innovation in their work.⁴ The subgroup is open to all, whether you consider yourself an AI subject matter expert or a complete novice who is simply interested in learning more about this field and its actuarial applications.

What is a subgroup?

The AI subgroup is modeled after the highly successful Health Community subgroups which focus on specific aspects of the healthcare industry (e.g., Medicaid, Pharmacy, Value-Based Care). One of the primary functions of the subgroup is to hold regular calls with members to share information and discuss topical items in the ever-evolving field of AI. The SOA is full of brilliant people doing exciting work, and these calls serve as a great opportunity to meet, engage with, and learn from others with shared professional interests.

Tell me more about the calls!

The AI subgroup calls are scheduled monthly, generally the first Thursday of each month from 11am to 12pm CT. Occasionally we may reschedule due to holidays, speaker conflicts, or other considerations. While there is typically a volunteer presenter (or two) for each call, the format is intended to be an open forum. The speakers help focus the conversation, but audience participation is strongly encouraged. While some materials may be distributed before or after each call, please note that calls are not recorded, so be sure to join live!

What topics have been covered recently?

We just began hosting these calls in June. The kickoff call served to introduce the subgroup and included a recap/discussion of the second AI Insights for Actuaries symposium held in May. Other calls have covered topics such as AI courses available on PD Edge+, vibe coding basics and pros/cons, building Excel models with LLMs

³ You can find information about the Emerging Topics Community and how you can get involved here <https://www.soa.org/communities/emerging-topics/>.

⁴ You can find more information about the SOA's strategic plan here <https://www.soa.org/about/strategic-plan/>.

(specifically Microsoft Copilot), and emerging AI regulations in the European Union. There is a wide range of topics, and we are always open to future suggestions (as well as volunteer speakers).

How do I get involved?

The first step is to sign up for the SOA’s **Emerging Topics—Artificial Intelligence Listserv**.⁵ Note that this is different from the **Artificial Intelligence Research Listserv** (although if you see other subgroups you’d like to join, feel free to sign up for more). Once your request is processed, you will receive an invitation to the recurring meeting series. Join the calls as your schedule allows and participate in the discussions. That’s it! If you’d like to take things a step further, you can also volunteer to speak on a relevant topic that interests you. This is a welcoming, low-pressure environment, and even if you only have 5–10 minutes of content, we’re always interested in hearing from other members of the community.

I hope to see you at the next call!

Andrew Dilworth FSA, MAAA is Senior Director, Actuarial Services at UnitedHealthcare in Appleton, Wisconsin

ACTUARIAL INTELLIGENCE BULLETIN

Is This the Last Invention or the Next Valuation?

RONALD POON-AFFAT, FSA, FIA, MAAA, CFA, HIBA

Isaac Saul, the American journalist who founded the bipartisan newsletter *Tangle*, recently published a provocative essay titled “*Is This the Last Invention?*”—a deep dive into the global debate over artificial intelligence.

Saul’s article divides today’s AI debate into three philosophical camps—the Doomers, the Accelerationists, and the Scouts. Doomers fear AI’s existential risks and want to halt its progress, Accelerationists champion rapid advancement to unlock its benefits, and Scouts accept AI’s inevitability while preparing society to adapt responsibly.

Actuaries may recognize echoes of these same instincts—prudence, curiosity, and accountability—across the broad spectrum of their daily work. None of this is absolute, of course, but acknowledging these tendencies helps reveal how each of us engages with emerging tools and navigates the challenge of change itself.

The real question is not whether the system works, but who ensures that it works fairly, prudently, and within bounds that protect both the institution and the public it serves.

Each scenario leads to the same conclusion: the actuary remains the human in the loop—the interpreter of intelligence, the custodian of accountability, and the guardian of trust.

⁵ Click “JOIN” immediately under “Artificial Intelligence Research Listserv” at <https://www.soa.org/news-and-publications/listservs/list-public-listservs/>

This article considers how these contrasting mindsets, when balanced, can shape a more resilient and forward-looking actuarial approach to AI.

The Doomers

(Representative: Eliezer Yudkowsky, AI-safety researcher known for warning about unaligned super-intelligence)

Doomers argue that AI development must be slowed—even paused—until society can guarantee control. In actuarial terms, one might imagine valuation or financial-reporting teams leaning this way: they prize reproducibility, documentation, and sign-off before anything new touches the balance sheet.

For them, “black-box modeling” isn’t progress; it’s a pending audit finding. Their unofficial motto echoes the 18th-century Scottish philosopher David Hume, who cautioned that “a wise man proportions his belief to the evidence”—the Enlightenment’s elegant way of saying, *if I can’t verify it, it doesn’t exist*.

They aren’t Luddites. They’re risk managers who see transparency as the ultimate safeguard against chaos.

The Accelerationists

(Representative: Elon Musk, entrepreneur emblematic of the “move fast and build” ethos)

Accelerationists see AI as both inevitable and transformative. They recognize that the actuarial process—once bound by quarterly cycles and data bottlenecks—is being rebuilt in real time. Product and analytics teams are deploying AI to reconcile data streams, generate alternative pricing paths, and deliver insights directly to decision-makers.

They argue that AI doesn’t eliminate actuarial oversight—it redefines it. By automating mechanical steps, actuaries gain bandwidth for interpretation, governance, and communication. In this model, speed becomes a form of strategy: rapid iteration produces resilience, and real-time learning drives relevance.

Their credo might be summed up in Mark Twain’s words: *“Continuous improvement is better than delayed perfection.”* Each iteration sharpens judgment, not just code—or as we might say today, *don’t slow down, fine-tune*.

The Scouts

(Representative: Geoffrey Hinton, the British Canadian computer scientist often called the “godfather of deep learning”)

Scouts stand between panic and euphoria. Like Hinton—who helped invent modern neural networks and later urged caution about their power—they believe AI can’t be stopped, only stewarded.

In actuarial life, risk-management, reinsurance, or investment teams may resemble this group. They run reinforcement-learning simulations yet insist on human override triggers; they let language models draft treaty clauses but run fairness diagnostics before acceptance.

Their creed borrows from Ronald Reagan’s Cold War wisdom: *“Trust but verify.”* They translate innovation into accountability.

How These Tribes Map (Imperfectly) to Actuarial Work

Actuarial Function	Possible AI Mindset	Illustrative Example
Valuation & Reporting	Doomer-leaning	Prefers transparent, explainable models for IFRS 17 roll-forwards before approval.
Pricing & Product Design	Accelerationist-leaning	Experiments with generative AI to explore new micro-insurance demand curves.
Risk & Capital Management	Scout-leaning	Uses AI for stress testing but keeps manual oversight logs.
Claims & Operations	Accelerationist / Scout mix	Employs LLMs for claim summaries while monitoring bias metrics.
Investments & ALM	Scout-leaning	Tests reinforcement-learning strategies under human governance.

AI doesn't distinguish between SOA and CAS—it amplifies how you think. The same model that unsettles one actuary may empower another.

From Calculation to the Interpretation of AI: The Actuary's New Frontier

In each scenario, the actuary remains the human in the loop: the interpreter of intelligence, the custodian of accountability, and the guardian of trust.

Imagine, for instance, a claims algorithm that denies coverage without clear justification—who verifies fairness? A dynamic annuity that recalculates its value every hour—who validates solvency? Or a reinsurer's language model that drafts treaty clauses faster than any human could—who ensures compliance?

In each case, the real question is not whether the system works, but who ensures that it works fairly, prudently, and within bounds that protect both the institution and the public it serves.

Each scenario leads to the same conclusion: the actuary remains the human in the loop—the interpreter of intelligence, the custodian of accountability, and the guardian of trust.

The Evolution of a Risk Ecosystem

This is why AI isn't our last invention, but the next chapter of actuarial imagination—an age where model risk expands to include ethical risk, making us answerable for the societal impact of our algorithms.

Across history, the idea of a trinity has symbolized balance and interdependence—three distinct forces held in creative tension. In the New Testament, the Holy Trinity represents unity of purpose among Father, Son, and Spirit. In political thought, the division of government into legislative, executive, and judicial branches was designed to achieve the same equilibrium through checks and balances. Even in classical physics, Newton's three laws of motion describe a self-regulating system where force, motion, and reaction coexist in balance.

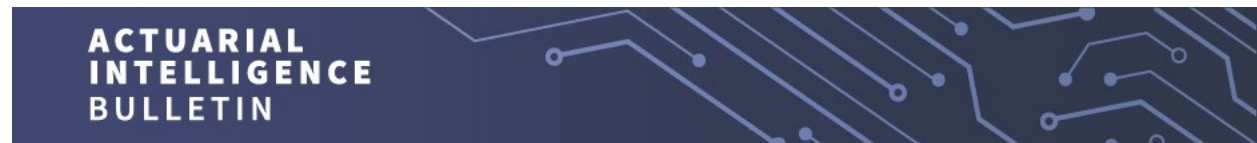
Rather than provoking rivalry or reinforcing divisions, the coexistence of differing actuarial outlooks—the Actuarial Doomers, the Accelerationists, and the Scouts—should foster mutual respect, for this triad, when aligned, reveals the profession at its most complete. In that same spirit, I envision the actuarial field evolving into a constellation of complementary forces, where caution, ambition, and stewardship sustain a pragmatic ecosystem in equilibrium—one capable of harnessing the very best that AI has to offer.

And there has never been a better time to be an actuary: part data scientist, part risk manager, part financial engineer, and part steward of responsibility—watching the greatest experiment in human reasoning unfold before our eyes.

Ronald Poon-Affat, FSA, FIA, MAAA, CFA, HIBA, is an independent board director, cross-continental actuary, and senior consultant.
Contact: rpoonaffat@soa.org

Isaac Saul's full article in Tangle can be read here:

<https://www.readtangle.com/the-last-invention-longview-artificial-intelligence/?ref=tangle-newsletter>



How to Insure AI? AI Insurance is Already Here

FEDERICO TASSARA, ASA

In recent years, we have witnessed numerous applications of artificial intelligence (AI) in the insurance industry, ranging from claims settlement and fraud prevention to underwriting policies. A multidisciplinary team, based in the United States and Germany, is working on something totally new. It's not AI for insurance; it's insurance for AI.

The Challenges

According to Gartner, 86% of CEOs believe that AI can grow their results in the short term, and at the same time, EY reports that 60% of executives are reluctant to apply it due to uncertainty about the profitability of these implementation projects.

Even the best-trained AI models will make mistakes, due to the probabilistic nature of their predictions. And any errors that AI commits are systematic in nature, meaning it will continue to make these mistakes until they are identified and corrected. However, the question as to how existing insurance policies protect insureds if AI goes wrong is still unanswered.

Currently, traditional insurance policies generally do not explicitly address artificial intelligence risks by way of affirmative covers or specific exclusions. Certain traditional lines of business may provide coverage for AI-related losses which are embedded in existing exposures and could lead to financial losses, bodily or personal injury or property damage and business interruption, like an employment practices liability insurance (EPLI) policy covering discrimination in hiring processes, or professional indemnity insurance covering potential errors stemming from AI-enabled professional services.

Technology errors and omissions (Tech E&O) insurance can cover AI vendors, but might not be enough to comprehensively respond to AI risk scenarios because, among other reasons, contractual liabilities are often excluded or limited, and triggers require a failure of the technology to meet a particular standard—but such standard is likely not near 100%, leaving a gap of un(der)covered losses for customers.

Today, selected players in the insurance market have started to discuss or even launch exclusions in policies for AI-related losses. Conversely, affirmative coverage for AI risks is still sparse, but does exist.

The Solutions

An insurance product was designed to address the unique risks and challenges associated with AI technologies, and covers financial losses or liabilities caused by errors or hallucinations of an AI or GenAI solution. It has different types of applications:

- **AI vendors.** Selling innovation can be challenging when customers lack trust in AI. Providing a performance guarantee for the AI models accuracy de-risks the purchase decision, but it creates a liability on the balance sheet of the tech company. Incorporating this coverage allows AI solution providers to provide a differentiating guarantee to their customers without the need to assume that liability. It gives them the possibility to clear the doubts that naturally arise in customers who are considering incorporating them as an AI solution somewhere in their value chain.
- **Enterprises using AI.** Rolling out AI models to automate critical business processes has huge upside potential but comes with uncertainty. AI errors can lead to ill-informed decisions, business interruption or other forms of lost revenue. Having coverage for these deviations from what was expected brings reliability and predictability to implementations and gives the company confidence to move forward in innovating safely.

Some of the success stories where this type of solution is already working are AI-enabled fraud detection, AI-powered alternative-asset valuations for assets used as collateral for loans, AI crop plans for farmers, and more.

The coverage for GenAI IP Infringement risks is also available to AI vendors (i.e., model providers), as well as Enterprises using AI (e.g., media-service companies, and corporations that frequently publish content, e.g., for advertising). The policy covers generated outputs to the extent that the users of the GenAI service did not intentionally create or use generated output to infringe on the rights of others. Coverage can include indemnification of legal fees and penalties as well as business interruption or re-work costs.

This comes at a time when some of the major GenAI providers have started offering indemnities for their (paying) customers. Anthropic, Amazon, Cohere, OpenAI, Google, Adobe, and Microsoft are some of the front-runners in this field and offer to pay back the costs that their customers incur if they face legal claims for copyright infringement from the use of their GenAI models—to varying degrees. This coverage can back those types of indemnities.

Our Role

The objective of insurance is to help society to be encouraged to do more, giving a framework of confidence to do the productive activities we do every day. From time to time, new technologies appear and with them new risks that need to be covered. That is the dynamic role that actuaries have in finding new ways to insure these new risks, challenging ourselves, and sharing knowledge with professionals from different disciplines to understand how new risks behave and from that how to insure them.

At some point in history, cars appeared, and with them the need for car insurance. The actuaries had the job of understanding the risk and creating the basis for a product that today is very common but at the time did not exist. The same thing happened with other technological advances of humanity. And today it is our turn with artificial intelligence. We are at the transcendental moment of understanding how it works to insure it. Projects like this one, where data scientists, economists, actuaries, lawyers, and other professions work together, are highly motivating and worth following closely.

Actuaries have a dynamic role: new technologies appear, bringing new risks for actuaries to find new ways to insure them.

Federico Tassara, ASA is Business Development Manager for Southern Europe and Latin America with Munich Re.

**ACTUARIAL
INTELLIGENCE
BULLETIN**

Artificial Intelligence Research

The Society of Actuaries Research Institute is building a network of AI expertise to inform, support, and strengthen the Actuarial profession.

From Health Inequities to Societal Bias: Insights from an Expert Panel on AI and Actuarial Responsibility

Ronald L. Poon Affat, FSA, FIA, MAAA

A panel of experts explore how potential bias in artificial intelligence is affecting the insurance industry—and how actuaries can evolve to meet this challenge.

<https://www.soa.org/resources/research-reports/2025/ai-actuarial-bias-equity>

The Impact of Artificial Intelligence on Retirement Professionals and Retirees

Michael Kummer, BA, Stefanos Orfanos, FSA, CERA, Niranjan Rajendran, B.Sc. (Hons)

Essays explore the impact of artificial intelligence (AI) and large language models (LLM) on retirement professionals and retirees.

<https://www.soa.org/resources/research-reports/2025/ai-retirement-essay-collection/>

ACTUARIAL INTELLIGENCE BULLETIN

Vibe Coding – Gutenberg’s Printing Press or Pandora’s Box

ANTON KOBELEV ASA, CERA, ENOCH OU, ALINA YE

Vibe coding is the term coined by AI expert Andrej Karpathy, founder of OpenAI, in February 2025. Vibe coding is essentially the process of using Large Language Models (LLMs), such as ChatGPT or any of the myriads of equivalent tools on the market right now, to describe the end-result in English and have the GPT generate working code for you.

The technology enthusiast and optimist in me is thrilled, but the prudent actuary in me is extremely cautious.

I am split on the concept of vibe coding. On the one hand, the technology enthusiast and optimist in me is thrilled by the prospect of getting things done without completing numerous coding courses or getting a degree in computer science. On the other hand, the prudent actuary in me is extremely cautious of the prospect of letting unsavvy actuaries loose and have them create tools that are erroneous or misleading.

Therefore, I will write this article in the voices of the two personas, battling in my head.

Andrej essentially defined what I have been doing for years—asking Google to find me a solution for a tiny spec of a larger problem, find some code on the web, adapt it as much as I can, integrate it into larger code base, and hope it compiles. I always felt handicapped by lacking in-depth coding skills, and my team was always empowered when an actuarial coder was among us.

Right now, I feel empowered myself. There are ways that AI can benefit actuaries in the realm of coding, especially on open-code systems like Python or R. The benefits include, but are not limited to:

- **Increased Speed of Model and Tool Development:** Vibe coding aims to significantly increase the speed of software development and the volume of customer-facing features developed. Actuaries could potentially leverage this by describing complex actuarial models, valuation systems, or data analysis tools to an AI using conceptual directions. The AI could then generate the underlying code, potentially much faster than traditional manual coding, accelerating the development cycle for new models or updates.
- **Focus on Actuarial Logic and Design:** Vibe coding shifts the focus to the overall product’s behavior and the “what” rather than the “how” of the code. This could allow actuaries to concentrate more on the critical actuarial logic, assumptions, and design of models and analyses, rather than getting bogged down in the specifics of programming syntax or debugging code line by line. They could describe the desired calculations, data transformations, or model structures, and the AI would handle the code generation.
- **Faster Prototyping and Scenario Testing:** Vibe coding could enable super-speed prototyping and iterative refinement. By describing variations conceptually to an AI, actuaries could rapidly generate and test the code for these different scenarios, iterating quickly based on the results.
- **Better Documentation:** Documentation and comments are always an after-thought and AI can generate proper documentation as the code is being created, as well as after it has been finalized. This will reduce the strain on new developers as they get rolled onto the project and must decipher someone else’s code.

The above points make a clear case that vibe coding using AI is a useful tool in any actuary’s arsenal. However, rapid and haphazard deployment of this approach is likely to be fraught with risk.

Let us now forget that current LLMs are essentially autocomplete engines on steroids that cannot reason, nor do they have intuition. These models are not computer scientists, who know about design, architecture, or flow of code. The solutions generated are most likely snippets of code found online, submitted by other users. These snippets are then assembled into a Frankenstein monster that is your model, based on your prompt.

The risk averse actuary in me is terrified and skeptical of having a myriad of mini models floating in my organization. There are four key concerns that make me lose sleep at night.

- **Verifiability:** Earlier this year, the Canadian regulator Office of the Superintendent of Financial Institutions released the guideline on Model Risk Management (E-23), which recommends ways financial institutions should approach model risk. A significant amount of space in the guideline is dedicated to model validation and defining what is a model or a mere tool. Other jurisdictions have similar guidelines that require users of models to validate them.

The ability to validate a model relies on two key principals—being able to interpret the model and having the know-how to replicate it or validate it in other ways. Models created using vibe coding are likely to be programs coded in the likes of Python or R, which require specialized knowledge from the validation team. These models are likely to be much more involved than Excel spreadsheets.

- **Maintainability:** Code can become a black box, with unclear logic and inconsistent naming conventions. AI can also often generate elegant solutions that are difficult to comprehend and thus difficult to enhance in the future. These AI solutions may also lack the modularity required for a production solution so that a certain component can be safely or easily updated without affecting other components.

Adjusting the vibe-code with prompt-after-prompt, may also affect the code’s architecture in a way that creates technical debt. Quick fixes and patchwork solutions accumulate without considering the big-picture design, making future development harder.

- **Scalability:** AI tends to prioritize correctness over efficiency, leading to slow or bloated applications. Vibe-coded apps might work fine for a few users, but struggle under real-world load.
- **Security:** AI-generated code sometimes skips input validation, making it prone to SQL injection, XSS, and other attacks. Vibe-coders might also accidentally hardcode credentials or insecure access patterns.

A rapid prototype is a fantastic springboard for discussions—a way to road-test ideas, frame assumptions, and get buy-in before diving deep. Using AI to vibe code such a prototype, the equivalent of a sketch on a napkin, is powerful and transformative, however do not discount the necessity of an SME in the room and the need to tread carefully. A human in the loop can be there to assure that the code is correct.

Anton Kobelev, ASA, CERA is the Executive Director at KPMG Canada’s Actuarial Modeling and Risk Management practice and is based in Toronto

Enoch Ou is the Manager at KPMG Canada’s Actuarial Modeling and Risk Management practice and is based in Toronto

Alina Ye is the Senior Consultant at KPMG Canada’s Actuarial Modeling and Risk Management practice and is based in Montreal

ACTUARIAL INTELLIGENCE BULLETIN

Creating and Using Synthetic Data

TOM CALLAHAN, FSA, MAAA, MS

Synthetic data is information that is artificially manufactured rather than generated by real-world events.⁶ It can include all kinds of data: images, tabular data, speech, music, and so forth. Synthetic data was once a quiet part of the data science universe, often considered in conjunction with oversampling techniques or when modeling fringe events.

However, the advancement of artificial intelligence has now caused an insatiable need for model training data. This need is so acute, and the competitive advantages so precious, that some companies are experimenting with using synthetic data to train their large language models.⁷ Therefore, synthetic data has the potential to become the next frontier in the artificial intelligence revolution.

This article will give an overview of ways to generate synthetic data, and it will touch on some use cases. It will focus on tabular data and sequential data, as these are the main data types that actuaries work with.

⁶ Hashemi-Pour, Yasar, and Laskowski. 2024 December 26. “What Is Synthetic Data? Examples, Use Cases and Benefits.” TechTarget.

<https://www.techtarget.com/searchcio/definition/synthetic-data>.

⁷ Seetharaman, D. 2024 April 1. “For Data-Guzzling AI Companies, the Internet Is Too Small.” The Wall Street Journal. https://www.wsj.com/tech/ai/ai-training-data-synthetic-openai-anthropic-9230f8d8?mod=article_inline.

Ways to Generate Tabular Synthetic Data⁸

A **rules-based approach** can be compared to “IF” or “RANDBETWEEN” formulas in Microsoft Excel. There is a random element to the creation of the data, but rules are established to constrain the output to a realistic range of values.

This approach relies on domain knowledge, and as such has the advantage of giving the programmer control over the output to ensure realism. However, for high-dimensional datasets, it could be unwieldy to develop realistic rules for every dimension while also ensuring that the dimensions properly correlate with each other.

A **copula** uses a combination of statistical distributions and correlations to generate synthetic data. The programmer defines each column with a statistical distribution and all the necessary parameters (for example, a column could be assigned as normally distributed with mean and standard deviation based on the underlying actual data). In addition, the user defines a copula function to ensure that the columns stay properly correlated with each other. The most common approach is to use the Gaussian Copula, which makes use of the data’s correlation matrix.

This approach works best when the desired outcome can be adequately described using statistical distributions.

Autoencoders are a special class of neural networks, in which the target output is the same as the input. This makes autoencoders useful for data compression, feature extraction, and anomaly detection.

The *variational* autoencoder produces synthetic data by introducing a random sampling element. Therefore, the output is targeted to be the same as the input—but the random element causes the result to be synthetic data.

Variational Autoencoders are useful when the main goal is to preserve an underlying “target” or “essence” from the input data.

Generative Adversarial Networks (GANs) are more common for synthetic image generation but can be used for tabular data.

GANs work by training a “generator” model and a “discriminator” model which compete against each other. (Hence the “adversarial” in the name.) The generator model creates fake data, and this fake data gets mixed in with actual data. This mixed data is fed to the discriminator model, which then guesses which data points are real and which are synthetic. The generator is trying to fool the discriminator; it is trying to generate fake data that the discriminator mislabels as real data.

The discriminator results get fed back to the generator model, which then adjusts its approach to do a better job fooling the discriminator. The loop continues until the discriminator cannot reliably distinguish real versus synthetic data.

Of the four approaches for tabular data in this article, the first two approaches (rules-based approaches and copulas) can help actuaries to control the process of the synthetic data generation process, whereas the last two approaches (autoencoders and GANs) are more black-box and require a lot of data, calibration, computing power, and review.

⁸ Much of this material was shared by Tom Callahan, Andrew Dilworth, and Joe Long at “Introduction to Synthetic Data,” on May 15, 2025, during the Society of Actuaries’ AI Insights for Actuaries Symposium.

Ways to Generate Sequential Synthetic Data

Long Short-Term Memory Models (LSTM) are a class of variant of recurrent neural networks (RNN) that process data sequentially. It works by retaining useful information from previous data points and then incorporating information from the latest data point.

LSTMs can be used to generate sequential synthetic data, such as text, time series, or tabular sequential data.

Transformers are the architecture underlying popular large language models such as ChatGPT. Rather than process data sequentially (as in an LSTM), transformers consider sequence as an attribute of the data. This change allows for more efficient processing and enables transformers to train on much larger datasets efficiently.

Applications for Synthetic Data

Some use cases of synthetic data include:

- Synthetic images of roads or construction, to understand driver behavior or forecast how traffic could change in response to changes in traffic patterns⁹
- Synthetic images of faces, to train facial recognition models
- Synthetic charts of electrocardiograms, to generate more plentiful data to support developing models to identify arrhythmias¹⁰
- It's been explored to generate synthetic health insurance claims and enrollment data.

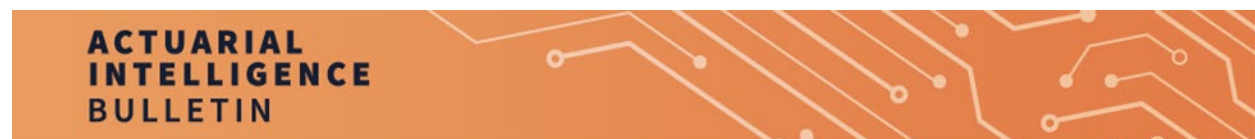
Some other potential uses include:

- Data security—it is safer to generate synthetic data as opposed to anonymizing actual data
- Data quality checks—if the synthetic data and actual data are very different, it could indicate a problem
- Data volume—especially useful, and less expensive, when considering tail events such as an economic crash or pandemic
- Data generation on demand

Conclusion

As artificial intelligence continues to grow in scope and impact, so will its need for data. At this point, the popular LLMs have largely absorbed the available data on the internet and in texts. It thus begs the question of how LLMs can be further enhanced, and synthetic data generation could be the answer. There are many risks involved with using synthetic data, to list a few, diluting true signals and distributional shift, so care must be taken to prevent its misuse. But, when combined with today's computing power and available technology, the creative possibilities are tantalizing.

Tom Callahan, FSA, MAAA, MS Manager & Actuary at Independence Blue Cross



⁹ Miaomiao, Y. 23–25 October 2024. "Synthetic Data Generation for AI Training." Blender Conference 2024 <https://conference.blender.org/2024/>, <https://www.youtube.com/watch?v=iuDv3TS-xQA>.

¹⁰ Clemente, F. 29 April 2020. "Synthetic Tabular Data Generation." Data Science Portugal; Meetup #76 Webinar. <https://www.youtube.com/watch?v=izMfWkWNtmA&t=10s>.

The Fifth Kind of Actuary: Expanding Our Thinking for a VUCA World

DAVID INGRAM FSA, CERA

Actuaries have always been committed to ensuring financial security, stability, and fairness in uncertain worlds. That mission has not changed. What has changed is the context: a world that is increasingly volatile, uncertain, complex, and ambiguous—and a world where artificial intelligence is beginning to take on much of the technical modeling work. In this setting, I want to suggest a possible new actuarial specialty. This “Fifth Kind” of actuary would build on the strengths of the first four kinds, while layering in additional ways of thinking. Its purpose is not to replace the traditional actuary, but to ensure that our longstanding objectives can still be achieved in today’s conditions.

VUCA, which stands for Volatility, Uncertainty, Complexity and Ambiguity is shorthand for a world teetering on the edge of out of control. The term was coined by the Pentagon in the early 1990’s when the geopolitical world moved into new territory with the break-up of the Soviet Union and the end of the bi-polar world. Each decade since has included one or several major shocks that have continued the streak of VUCA experiences for all of us. Plenty for the Fifth Kind of Actuary to worry about.

This idea follows naturally from two articles in the September issue of the AI Bulletin. In “A Future for Humans,” I explored how actuaries can assert their enduring value alongside AI. In “The Actuarial AI Competency Ladder,” the practical skills actuaries need to become effective AI collaborators—what I call the “Fourth Kind of actuary.” Those pieces were about grounding ourselves in the present. Here, I want to look further ahead.

My starting point comes from a framing offered years ago by CAS President Steve D’Arcy, who described the First, Second, and Third Kinds of actuaries: deterministic, stochastic, and financial-market-consistent. To that lineage we can now add the Fourth Kind (AI actuaries), and now a Fifth Kind—not yet widely present but emerging as a suggestion for where the profession might go in response to both AI’s new capabilities and the demands of a VUCA world.

Five Ways of Thinking

At the heart of this Fifth Kind is metacognition—the ability to step back, reflect, and intentionally select different modes of thinking depending on the challenge at hand. I suggest that actuaries in this specialty will combine five ways of thinking: our traditional actuarial discipline plus four additional layers.

- **Actuarial Thinking** emphasizes rigor, quantification, and discipline. It uses data, probability, and models to measure risk, set prices, and ensure solvency. Its strength lies in bringing clarity and structure to uncertainty, giving decision-makers a sound numerical foundation.
- **Systems Thinking** emphasizes interconnections, feedback loops, and unintended consequences. It shifts the focus from isolated risks to dynamic systems, showing how actions in one area ripple through others. It equips actuaries to anticipate cascades and design resilient strategies rather than static fixes.
- **Plural Rationalities Thinking** emphasizes contrasting worldviews and stakeholder legitimacy. It recognizes that what is “rational” for one group (solvency, affordability, fairness, political stability) may be irrational for another. By mapping and mediating these perspectives, actuaries can craft solutions that stand a chance of being implemented.

- **Ethical Reasoning** emphasizes fairness, justice, and trust. It asks who benefits, who bears the burden, and whether a proposal is legitimate in the eyes of the public. This ensures actuarial solutions are not only technically sound but also morally defensible and socially sustainable.
- **Creative Thinking** emphasizes imagination, innovation, and adaptability. It generates new products, structures, and approaches when traditional tools no longer work. Creative thinking allows actuaries to transform constraints into opportunities, designing solutions for risks that have no historical precedent.

What the Layers Add

Each layer brings something vital when added to actuarial thinking.

- Actuarial + Systems: moves us from static risk pricing to dynamic resilience planning.
- Actuarial + Plural Rationalities: prevents “technically correct but socially rejected” outcomes.
- Actuarial + Ethical: safeguards fairness and legitimacy alongside solvency.
- Actuarial + Creative: enables innovation in the face of unprecedented risks.

Together, these create a toolkit that allows actuaries to address problems that are otherwise intractable. Systems, Plural Rationalities, Ethics and Creativity are also all areas where AI struggles to operate.

Strategic Leadership and Wicked Problems

Where would such actuaries focus their energy? Two domains stand out: strategic leadership and wicked problems.

Strategic leadership: As AI assumes more of the routine work of modeling, actuaries who can interpret, integrate, and guide will become more valuable than those who only calculate. Fifth Kind actuaries will sit at decision-making tables—in C-suites, on boards, and in government task forces—translating risk into strategies that balance solvency, resilience, fairness, and legitimacy. Their edge will come from plural rationalities, emotional intelligence, and ethical reasoning.

Wicked problems: These are the messy, interconnected, contested challenges of the VUCA world—climate change, social security sustainability, health systems, cyber resilience. They cannot be “solved” once and for all; they can only be navigated. Fifth Kind actuaries will use systems thinking to map interdependencies, social sciences to reconcile stakeholder tensions, and creative thinking to design adaptive, resilient interventions.

A Specialty, Not a Replacement

It is important to be clear: this Fifth Kind does not diminish the first four. Deterministic, stochastic, financial-market-consistent, and AI-collaborative actuaries remain indispensable. The Fifth Kind should be seen as a specialty track—relevant for those working on systemic risks and strategic leadership but not required of all actuaries. In that respect, it resembles how financial-market-consistent modeling began as a specialty and later became mainstream. Time will tell whether the Fifth Kind follows that path, but for now it is best understood as a suggestion—a possible direction for actuaries who want to extend their impact.

Still an Actuary

The Fifth Kind must remain firmly grounded in the actuarial identity. This means:

- Retaining core actuarial thinking as the foundation—solvency, quantification of risk, and professional discipline.
- Operating under the same professional standards, codes of conduct, and regulatory expectations as all actuaries.

- Applying new ways of thinking specifically to actuarial domains such as insurance, pensions, systemic financial risks, and capital management.
- Ensuring that technical rigor and accountability remain central, even when engaging in systems mapping, ethical reasoning, or creative design.

Conclusion

Actuaries have always defined themselves as guardians of risk, uncertainty, and security. That identity remains. What is shifting is the context—a VUCA world where risks are more interconnected, contested, and unpredictable, and where AI is taking on much of the technical work. In this environment, some actuaries will choose to branch out, cultivating metacognition and layering systems, plural rationalities, ethical, and creative thinking onto their actuarial foundation.

The result is the Fifth Kind of actuary: a specialty that equips us to handle wicked problems and strategic leadership challenges, while staying true to our traditional mission. It is not the future for all actuaries, but it may be the future for those who want to remain central in addressing the most complex risks of our time.

David Ingram FSA, CERA is an editor of the AI Bulletin in New York.



ACTUARIAL INTELLIGENCE BULLETIN

Editorial Committee

Jon Forster, ASA

Dave Ingram, FSA

Ronald Poon Affat, FSA

Zhiyu (Frank) Quan, PhD, FSA

Darryl Wagner, FSA

Associate Editor

Jing Kai Ong, ASA

Thank you for reading the November 2025 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 Research-AIB@soa.org.

About the Society of Actuaries Research Institute

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

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

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

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

Society of Actuaries Research Institute
8770 W Bryn Mawr, Suite 1000
Chicago, IL 60631
www.SOA.org