ACTUARIAL INTELLIGENCE BULLETIN

SOA Research INSTITUTE

July 2026

Welcome to the July 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.

Actuarial Intelligence Research Bulletin Board

SOA RESEARCH INSTITUTE

Welcome to the Actuarial Intelligence Research Bulletin Board! From practical applications and emerging opportunities to governance, ethics, and risk considerations, this bulletin board highlights selected SOA Research Institute work designed to help actuaries understand how AI is shaping the future of our work. You can find all of the SOA Research Institute's work on AI here.

Actuarial Mindsets for Leading in the AI Era
A quick, practical roadmap for actuaries who want to lead as AI becomes a routine part of the workflow. It highlights the enduring mindsets that matter most: curiosity, disciplined thinking, clear problem framing, and integrating AI responsibly without losing sight of core professional obligations.

AI in Actuarial Work
A collection of eight essays exploring a variety of actual, day-to-day experiences of practicing actuaries as they navigate AI.

AI in Healthcare and Health Insurance
Roundtable insights from actuaries and health leaders on how generative AI is transforming healthcare and health insurance, with practical use cases plus governance, transparency, and accountability guidance.

The Impact of AI on Retirement Professionals and Retirees
This essay collection explores the multifaceted impacts of AI/LLMs on retirement professionals, retirees and those planning for retirement. It serves as a resource that will aid in the understanding of relevant issues while providing current considerations and potential future dynamics of AI and LLMs in this area.

Applying Fairness Frameworks to Long-Term Care Insurance: Actuarial Considerations for AI and a Pricing Example
Explore practical guidance to assess and improve fairness when using AI in long-term care insurance.

AI Research Spotlight

Experience Studies Harnessing an AI Agent–A Proof-of-Concept Lightyears Past Code Generation

Mark Spong's essay challenges the common view of AI as merely a coding or writing assistant, demonstrating how actuaries can collaborate with AI agents to perform end-to-end experience studies. Using a mortality dataset and a direct connection between AI and a Python environment, Mark shows how AI can autonomously explore data, generate and execute code, test hypotheses, build and evaluate models, attribute drivers of results, and draft actuarial reports—all while the actuary provides oversight and judgment. The proof of concept illustrates a practical workflow for asking business questions in natural language and guiding AI through complex analytical investigations rather than coding each step manually. It also highlights potential organizational hurdles, including data governance, auditability, workflow redesign, and cost management.

AI in the Actuarial C Suite: Maintaining Credibility

What Senior Leaders Might Want to Ask as AI Becomes Embedded in Actuarial Work

DAVE DILLON, FSA, MAAA

In "AI in the Actuarial C Suite: Cost and Capability" we talked about how to go about achieving new governance demands, model risk considerations, cybersecurity exposure, and client and regulator expectations. Here we will pivot to considerations of maintaining professional quality and reliability both in the short and long term.

Adapting the infrastructure to work better with AI is likely to be important, but will it be enough? That may depend whether actuaries on staff are willing and able to take advantage of the new tools to execute and to bring the bring the results to the executive team. That would require skills, especially communications skills, as well as a culture that supports using AI to get the work done, not just faster but also at the same level of quality. For the long run, the training pipeline for actuaries needs to be modified as well, emphasizing the development of different capabilities.

1. Communication and Stakeholder Expectations: AI Drafts, Humans Decide

As drafting and summarizing become easier, the "last mile" becomes more important: selecting what matters, clarifying uncertainty, and translating technical results into decision ready guidance. Executives will likely expect the bar to rise for communication quality, not fall.

These questions may become routine in executive, client, and regulatory settings:

- Are you using AI in this workstream, and if so, how?
- What data entered AI systems, and what controls prevented inappropriate disclosure?
- How did you validate AI assisted outputs, including any code or narrative conclusions?
- What human review occurred, and what documentation supports the final recommendation?

Organizations that can answer those questions will likely have an advantage. Those that cannot may face delays, distrust, or additional oversight.

2. Culture and Operating Model: Speed is Not the Only KPI

AI offers visible productivity, e.g., faster turnaround, more content, more output. But actuarial value often comes from preventing bad decisions through skepticism, controls, and well-communicated limitations. If incentives emphasize volume and velocity, AI can amplify those behaviors. If incentives emphasize rigor and defensibility, AI can amplify those instead.

A useful executive exercise may be to decide where the actuarial organization intends to compete:

- High throughput production (standardized deliverables, rapid cycles, consistent documentation).
- High trust advisory (scenario analysis, strategic insight, and decision support).

Most organizations will likely need elements of both for their AI and talent strategies to align with the desired operating model.

3. Talent Development: If AI Does More of the Drafting, How Do Actuaries Learn Judgment?

AI can compress early career work by drafting memos, outlining methods, and generating code. That can accelerate contribution, but it can also short circuit the apprenticeship model if not managed intentionally. The risk is a workforce that can prompt well but cannot defend assumptions under pressure.

Leaders can adapt development models in several practical ways:

- Teach actuarial judgment explicitly (assumption setting, credibility, reasonableness, tradeoffs).
- Require “curate and defend” deliverables, e.g., staff present results live and handle questions.
- Use controlled “AI off” repetitions for foundational skills early in training.
- Elevate the reviewer role and train reviewers, since review quality becomes the differentiator.
- Shift evaluation metrics away from speed alone toward clarity, rigor, and issue spotting

As AI drafts more, human judgment, communication, and accountability become the profession's competitive advantage.

4. Practical Next Steps

People are likely to remain an important part of business processes, especially for actuarial work. As AI capabilities evolve, it may be valuable to consider how actuarial skills, training and culture could adapt to changing needs and priorities. For example, consider:

- Redesigning training so staff build judgment, challenge and communication skills, not only prompt fluency.
- Measuring outcomes beyond "usage": error rates, rework, turnaround time, and stakeholder trust.

Closing Thought

I believe that AI will not replace actuarial judgment, but it will replace some traditional actuarial tasks. That would make AI a leadership test more than a technology test. Organizations that treat AI as a shortcut may risk quality failures and reputational damage. Organizations that treat AI as a disciplined capability, supported by governance, training, and accountability, may be able to elevate the actuarial function from production engine to strategic partner.

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

ACTUARIAL INTELLIGENCE BULLETIN

Monitoring of Deployed AI Systems Overview

DALE HALL FSA, MAAA

Actuaries spend a lot of time and effort developing and updating models. Our profession laboriously checks and rechecks model parameters, ensuring good fit and explainability both at implementation and over time. Actuarial capabilities around model development and validation have direct relevance and applicability to the emerging practice of building new models and systems that incorporate artificial intelligence (AI) and adhere to a new and

growing world of AI risk management and governance frameworks. This article explores the challenges and opportunities facing actuaries as they develop and work with models increasingly enabled by AI.

The Society of Actuaries Research Institute, as part of our involvement with the Center for AI Standards and Innovation (CASAI) at the National Institute of Standards and Technology (NIST), has been reviewing and discussing these challenges, especially with a focus of how AI is evolving across the insurance and financial services industries. A new NIST report entitled *Challenges to the Monitoring of Deployed AI Systems* highlights a critical shift in how organizations must think about artificial intelligence: not as a one-time model development exercise, but as an ongoing lifecycle requiring continuous evaluation, monitoring, and refinement.[1] While much of the AI field has historically focused on pre-deployment testing, the report emphasizes that real-world performance often diverges from controlled testing environments, making post-deployment monitoring essential to ensuring reliability, safety, and effectiveness.

Actuarial expertise in model validation and experience studies naturally extends to AI monitoring.

A central idea is that AI systems are inherently dynamic. Once deployed, they interact with changing data, evolving user behavior, and shifting environments. As a result, models can experience performance degradation, unintended behaviors, or "drift" over time. Monitoring is not simply about validation. It is increasingly more about detecting when a system no longer behaves as expected and needs updates, recalibration, or retraining. This aligns closely with actuarial practice, where models are periodically refreshed as experience emerges and assumptions are revisited.

The NIST Monitoring Framework

To structure this complex problem, NIST identifies six categories of post-deployment monitoring: functionality, operational performance, human factors, security, compliance, and large-scale impacts. These categories collectively recognize that AI risk extends beyond predictive accuracy. For example, a model may perform well technically but fail in terms of user interaction, regulatory compliance, or broader societal outcomes. This view mirrors our profession's familiar enterprise risk management approaches, where risks are assessed across operational, strategic, and reputational dimensions rather than viewed in isolation.

A key contribution of the report is its framework for understanding *gaps* and *barriers* in current monitoring practices. Gaps are defined as underexplored or insufficiently developed areas, while barriers are practical obstacles that hinder effective monitoring. This distinction is particularly useful for practitioners as gaps point to where more innovation and research are needed, while barriers highlight constraints that must be managed in implementation.

Among the most significant gaps is the lack of standardized methods and metrics for monitoring AI systems. Organizations often struggle to determine what should be measured and how to interpret results. This is compounded by the fact that monitoring is highly dependent on the context in which the model has been implemented. Metrics that are appropriate in one application may not generalize to another. The absence of widely accepted standards creates inconsistency and limits comparability across systems.

[1] Anita Rao, Andrew Keller, Neha Kalra, Ryan Steed, Kweku Kwegyir-Aggrey, Kevin Klyman, Diane Staheli, and Amanda Bergman, *Challenges to the Monitoring of Deployed AI Systems: Center for AI Standards and Innovation*, NIST AI 800-4 (Gaithersburg, MD: National Institute of Standards and Technology, March 2026), https://doi.org/10.6028/NIST.AI.800-4.

Barriers further complicate this landscape. The report identifies several cross-cutting challenges, including limited visibility into model behavior, insufficient information sharing across stakeholders, and the rapid pace of technological change. For example, organizations may not have access to the underlying data or model components needed to fully understand system performance, particularly when relying on third-party AI providers. At the same time, the speed at which AI systems evolve makes it difficult to maintain up-to-date monitoring frameworks or documentation.

Resource constraints also play a major role. Effective monitoring can require significant investment in data infrastructure, computational resources, and skilled personnel. Human-in-the-loop processes, while valuable for validation, can be difficult to scale as systems grow. Often, organizations spend a lot of time and resources on initial implementation but don't fully account for the resources needed for effective monitoring. These challenges create tension between the need for rigorous oversight and the practical realities of implementation.

The NIST Monitoring Framework expands AI oversight beyond model accuracy to six dimensions: functionality, operational performance, human factors, security, compliance, and societal impacts.

Challenges and Questions

At a more granular level, the report highlights category-specific issues that are especially relevant to model lifecycle management. In functionality monitoring, for instance, organizations face difficulties in establishing performance baselines and detecting drift over time. This is directly analogous to actuarial model monitoring, where identifying when experience deviates materially from expectations is a core task. Similarly, the lack of high-quality "ground truth" data in live environments makes it harder to assess model accuracy post-deployment.

Human factors monitoring presents particularly significant gaps. There is limited understanding of how users interact with AI systems, how their behavior changes over time, and how feedback loops between users and models might end up influencing outcomes. AI systems do not operate in isolation but are built, developed, and implemented in ways where human behavior is both an input and an outcome.

The report also raises a set of open questions that organizations need to resolve, including who is responsible for monitoring, what should be monitored, and how frequently monitoring should occur. These questions underscore the lack of consensus and point to the need that governance frameworks that clearly define roles, responsibilities, and processes.

Overall, the report reinforces the idea that post-deployment monitoring is not to be discounted and is a central component of creating trustworthy AI systems and models. It creates a feedback loop in which insights from real-world performance inform model updates, improve pre-deployment testing, and ultimately enhance system reliability.

Closing Observations for Actuarial and Risk Management Practice

First, this report provides a valuable conceptual bridge between traditional actuarial model governance and modern AI systems. Actuaries are already accustomed to model validation, experience monitoring, and periodic assumption updates. The NIST framework extends this idea into a broader environment, highlighting the important role actuaries can in ongoing future monitoring of the systems they construct.

Second, the emphasis on gaps and barriers highlights an opportunity for the actuarial profession to contribute meaningfully to AI risk management. Actuaries bring expertise in measurement under uncertainty, model lifecycle management, and risk-based decision-making—skills that are directly applicable to addressing the challenges

identified in the report. By adapting these capabilities to AI monitoring, insurers can strengthen governance, improve model reliability, and better manage emerging risks in an increasingly AI-driven landscape.

Dale Hall FSA, MAAA is Managing Director of the Society of Actuaries Research Institute

"We Can't Afford All These Humans:" A Doctor's Case for AI in Medicine

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

A review of Robert M. Wachter's A Giant Leap: How AI Is Transforming Healthcare and What That Means for Our Future *(with reference to his NPR interview of the same name)*

Picture two rooms, side by side, sometime in the future.

In the first, a large open space bathed in cool blue-white light. Touch screens and digital kiosks stand at intervals across the floor, each one occupied. A mother crouches beside her young daughter, guiding small hands across a glowing interface that is already asking about symptoms. Nearby, an elderly man sits alone on a bench, squinting at a screen, working through a triage sequence at his own pace. A young professional stands, barely looking up, navigating his way through what might once have been a fifteen-minute appointment — done now in three, on his feet, between meetings. There is no receptionist. There is no waiting room in the traditional sense. The room hums with quiet activity, people moving through their healthcare the way they move through a self-service terminal at an airport. Efficient. Frictionless. Available to anyone who walks in.

In the second room, everything slows down. The lighting is warm and low. There are bookshelves, a plant in the corner, a painting on the wall. Two chairs face each other at a comfortable angle, and in them sit a silver-haired physician and his patient — a woman, perhaps in her forties, leaning slightly forward. He is not typing. He is not glancing at a screen. He is listening, the way doctors once routinely listened, with time and full attention. The conversation appears to have been going on for a while. It looks like it might continue. This room costs something. It feels like it.

The two rooms are not in conflict. They may simply represent two tiers of the same near-future health system: one for the many, one for those who can afford something more. Whether that future represents a triumph of access or a quiet erosion of equity depends on how we build toward it. Robert Wachter's *A Giant Leap* is, among other things, an attempt to answer that question before the answer is made for us.

Wachter is a respected voice in modern American medicine. As Professor and Chair of the Department of Medicine at the University of California, San Franciso (UCSF) — and the physician widely credited with coining the term "hospitalist" — he brings a clinician's hard-won skepticism to every technology claim. His previous book, *The Digital Doctor*, documented the messy, disappointing first wave of healthcare digitization. *A Giant Leap* picks up where that story left off, asking whether AI might finally deliver what EHRs promised and failed to provide. NPR borrowed the book's most provocative line for a recent interview with Wachter, titling it "We Can't Afford All These Humans." It is an uncomfortable phrase — and a deliberately chosen one. Wachter is not predicting mass displacement. He is describing the fiscal reality already forcing the question in hospitals, insurers, and health systems: the current model of delivering care through an ever-expanding army of expensive clinicians and administrators is not sustainable. AI

may not be a choice so much as an arithmetic inevitability. The book, and the interview, explore what that means in practice.

This is not another breathless manifesto. It is more valuable than that.

The Watson Warning

IBM launched Watson Health in 2015, eventually investing roughly $4 billion in the ambition of making a machine that could think like a doctor. The assets were later sold at a fraction of that cost. Wachter uses this failure not as a cautionary tale against AI, but as a lesson in sequencing. IBM had already proved that machines could dominate bounded contests — Deep Blue beat Garry Kasparov in 1997, Watson won Jeopardy in 2011. But medicine was never chess or a game show. It was messier, more human, and far less forgiving. The smarter first move, Wachter suggests, might have been to build a brilliant medical secretary before attempting a brilliant physician.

Where AI Is Actually Winning

The first real killer application in medicine, Wachter argues, is not dramatic diagnosis. It is paperwork relief — notes, inbox messages, prior authorizations, and the administrative burden that drives clinician burnout. That sounds mundane. It is not. If AI gives physicians more time to think, listen, and treat, the boring application may turn out to be the transformational one. Readers of this bulletin will recognize the parallel in actuarial work, where documentation, communication, and summarization tasks are among the most frequently cited early wins.

> Paperwork relief—not diagnosis—may prove to be AI's first transformative contribution to healthcare.

Wachter's preferred frame for AI in clinical practice is the corridor consultation: a smart colleague you can ask what you might be missing, who can challenge your thinking and suggest alternatives, but who does not own the final judgment. That responsibility remains with the human. The analogy translates directly to actuarial settings where AI-assisted analysis still requires professional sign-off, documented reasoning, and defensible assumptions.

The Uncomfortable Implications

Two of Wachter's more pointed observations deserve attention. First, AI could eventually expose performance variation across physicians — comparing decisions, treatment patterns, missed diagnoses, and outcomes in ways that are data-driven and continuous. Doctors are accustomed to giving opinions. They may be less comfortable receiving systematic feedback on whether their decisions are better or worse than their peers. Analogous dynamics will likely emerge in actuarial model governance as AI benchmarking matures.

> The greatest value of AI may be giving professionals more time to think, listen, and exercise judgment.

Second, Wachter raises the prospect of AI-first health plans, where AI handles triage, monitoring, follow-up, and routine guidance, with access to a human physician repositioned as an upgrade tier. This is the second room made policy. The economics are straightforward, and insurers, employers, and governments will be drawn to it. Whether such a structure serves patients well is a separate question — and one Wachter does not shy away from.

Conclusion

The real lesson of *A Giant Leap* — and the honest answer embedded in that NPR headline — is that medical AI is unlikely to arrive as a robotic super-doctor. It will more likely arrive as an assistant, a note-taker, a second opinion, a feedback engine, and eventually a gatekeeper. Those two rooms at the start of this piece are not science fiction. The

infrastructure for them is being built now, quietly, in the choices that health systems, insurers, and policymakers are making today. For those in actuarial and risk disciplines thinking carefully about where AI creates durable value versus where it introduces new accountability burdens, this book is a useful and grounded place to start.

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

ACTUARIAL INTELLIGENCE BULLETIN

The Day We Learned What "AI" Means

MARIO DICARO, FCAS, CERA, MAAA

Someday, when the robots have taken over, we will celebrate November 30, 2022, as AI Day. That is the official public release date for ChatGPT. That was the day the general public realized that teams of engineers all over the world had (kind of) figured out how to get computers to engage in language-like conversation with humans, themselves, and other computers. ChatGPT quickly became the fastest growing app ever.

AI has had a huge impact on my professional and personal life. Prior to 2023 I would occasionally end up in some version of a conversation where my colleagues were trying to understand the difference in meaning between three phrases: "AI", "Machine Learning", and "Predictive Analytics." I promise this is relevant. This one repeated conversation reveals some subtle things about language we don't pay much attention to:

- We use language to encode and transfer knowledge
- Individuals encode their personal knowledge in language to improve their own recall. Think of journals and notes. Writing doesn't just contribute to recall, but also to formulation and structure of ideas.
- Groups use language in similar ways but also to get everyone thinking and working on the same thing. The words are created or modified to suit the group.
- Language evolves like fractals in nature.

A true anthropologist has a much better list of what language is used for, but that's all we need for this writeup. Today the internet defines those three phrases like this:

- AI is the outermost, largest category,
- ML is the technology driving the system, and
- Predictive Analytics is the specific outcome or business tool.

That's broadly what the internet has settled on as of June 2026. Three years ago, it was not so clear cut, or the conversation would not have resurfaced as often as it did.

November 30, 2022, the day that AI becomes real.

I lived through similar evolutions of vocabulary when Solvency II regulations rolled out in Europe, and more recently when my company underwent a multi-year exercise to do financial reporting under IFRS. Whenever large transformations like this happen, people must work together on those projects. People use words to communicate with each

other to get work done, and when the work is new, they need new words. A working vocabulary evolves among those doing the work. Maybe the group is only 10 people and the vocabulary settles quickly. But then a team from one company interacts with a team from another company. They may talk about "risk appetite" only to realize that that phrase has a slightly different meaning in different contexts. Or, even worse, people are using the phrase with no clear meaning—just using words because they happen to be in a group that uses those words.

In May of 2023 I attended the Generative AI Summit in London with a colleague of mine. It was a small conference, but very eye-opening. That conference introduced me to the types of problems LLMs were good at, gave me a vocabulary, and inspired me to get my hands dirty. I have been applying LLMs to our business problems ever since.

It's interesting that the internet-official AI/ML/PA breakout above does not define the word AI as it is commonly used today. **AI means: agentic systems using LLMs as brains.** In my opinion, November 30, 2022, was the day that we learned what the term "AI" meant because they finally saw in real life the image that the phrase conjures up in a mind. People intuitively know that language is fundamental to humanity and intelligence. "Artificial intelligence" must involve a machine (it's artificial) and that machine must be able to talk with a human (it's intelligent). Arms, legs, torso, wheels, electric motors, and monitors can all be present on a robot. Whether that robot is a car or a humanoid, until it talks, it is not AI—at least that's how I see it.

For the past couple of years, I've served as a volunteer in the CAS as the AI Working Group chair. It has been quite a bit of work, and a lot of fun. A question that comes up in nearly every volunteer or presentation-like setting is: How can one learn AI? I have a strong opinion on this topic, and it's simple: you learn AI skills the same way you learn any other skill.

Imagine an activity you are skilled at. Guitar, baseball, tae-kwon-do, running, mathematics, or writing, all serve well for this example. How did you get skilled? The process is always the same: study and practice. Either effort alone is unlikely to result in your being considered skilled at the activity. If you only practice, you will achieve some level of mediocrity, like the state of my jogging. If you only study, you could become very knowledgeable about the activity, but you wouldn't be considered skilled at it. If you were able to imagine an activity you are skilled at when I brought it up, I bet you've applied the same study + practice = skill formula to get where you are. The study tells you what to work on while you practice. The combination of study and practice moves you out to the edges of abilities of the human population. **This is where your abilities add value to the group.** This is how AI skills work. You identify a problem to practice on, and a body of knowledge to study, then you iterate over the cycle of study → practice → evaluate → study → practice → evaluate… and eventually your peers recognize that you are skilled.

To Learn AI:

STUDY + PRACTICE = SKILL

Mario DiCaro. MAAA FCAS CERA is VP Capital Modeling and Analytics at HCC Holdings Inc.

When the Trenches Collapse: Building the Next Generation of Actuarial Judgment

AREE BLY

As an analyst, it was common to hit "run" on an Excel macro or execute an SQL query only to have it crash or come back with nonsensical results. Often, it was a model that had worked fine for prior products or with a prior dataset. After a few hours of banging my head against the wall, I would come to a realization.

It was time to go back to first principles. Stop making the prior model fit the new challenge and go back to thinking about the problem we're trying to solve.

Sometimes that meant scratching the model entirely. Sometimes it meant using it in a fundamentally different way.

But it always meant releasing the assumption that tweaking the old process was best.

When you went back to first principles as an analyst, you were deep in the trenches. You were learning to think critically and building the judgment that would help you make faster and better decisions later in your career.

AI is collapsing the trenches where actuarial judgment has traditionally been built.

Those trenches weren't just where the work happened. They were ground zero where actuaries were formed.

AI is collapsing those trenches. The granular, iterative, sometimes frustrating work where so much of our intuition and judgment is built is changing.

Organizations are trying to figure out what that means. With such a potentially big leap to make, just tweaking how we develop actuaries may not be enough. We may need to go back to the fundamental question: what are we trying to solve?

What Senior Actuaries Are Up Against

AI is absorbing the granular work that used to build actuarial judgment.

At the same time, with the help of AI, junior actuaries are producing more output faster. But there is a tradeoff involved. While technically sound, it may not be actuarially sound.

This puts more pressure on senior actuaries who have extensive judgment and experience to catch issues.

And senior leaders don't have a map for this. AI-augmented actuarial work is new territory for everyone.

The good news is that senior actuaries, of all professionals, are well-positioned to assess what poor AI integration in an organization may cost and to build the business case for intentional judgment development as part of the actuarial job.

The question isn't whether to integrate AI, but how to do it well. Organizations may need to step off the old path and create a new path to judgment and experience.

As AI reshapes traditional learning paths, organizations may need to find new ways to cultivate actuarial judgment.

Four Things Senior Leaders Can Do

1. Consider assessing where your team is today.

   Many junior actuaries may already be using AI in different contexts. Senior leaders can ask them to walk through not just what the result is, but how they evaluated it, what they questioned, and what they're uncertain about. Strong judgment layered on AI use can be fine. But thin responses on polished output may highlight a gap to backfill.

2. Consider defining what first principles mean for your organization in an AI world.

   The first principles of the actuarial profession may include: models are tools and develop estimates; professional judgment cannot be outsourced; actuaries produce defensible numbers with understood limitations. But what those principles mean in practice for your market, your products, and your actuaries is something each organization needs to work out explicitly.

   Bringing people together across levels to define what actuarial judgment looks like when AI is part of the process may be helpful. What do human actuaries need to own? What does it mean to sign off on AI-assisted work? What does peer review of AI output look like?

3. Consider deliberately developing judgment.

   When judgment no longer naturally develops as a byproduct of doing the work, consider intentionally developing it. Here are a few ideas for what this could look like:

   - Pair junior actuaries with senior ones to observe judgment in action. Senior leaders can share what are normally internal thought processes, decision making, and gut checks. What prompts a second look? What question does the senior actuary ask before accepting a result? Make the judgment process visible.
   - Offer explicit training that builds critical thinking, decision making, and judgment. Encourage junior actuaries to articulate reasoning and thought processes to others involved in a project.
   - Regularly measure how your actuaries are developing their skills. For any work that they have done using AI augmented processes, ensure that they can defend it, extend it without going back to AI, and translate it to another context.

4. Consider working as equals in new territory.

   As organizations navigate AI-augmented work, it may be helpful for senior leaders to step out of the instructor role and assemble groups with actuaries at all levels to address specific challenges. Instead of putting the most senior person in the lead role, someone in the middle may be more effective at drawing out ideas and perspectives from all at the table as you map out the future.

   Co-creating solutions and new processes — working through problems side by side — may help to incorporate perspectives from all levels and lead to a more effective result. This approach would build the future of the actuarial path together, rather than handing down from senior to junior.

What This Builds

When organizations do these things consistently, teams could develop a shared view of what good actuarial judgment looks like in an AI-augmented world. They could build collective knowledge about where AI belongs and

where actuarial insights should remain human. They could grow a generation of actuaries who don't just use AI efficiently, but who know how to develop and use judgment.

That's the first principles answer: be willing to go back to ground zero together.

Aree Bly is Professional Trainer and Coach at Alignment Ally, LLC, and a former consulting actuary

When the Algorithm Answers for Itself: Colorado's AI Governance Framework for Insurance

SYED RAZA FSA, MAAA

An algorithm does not need to intend to discriminate in order to produce discriminatory outcomes. It only needs to be trained on data that reflects existing inequalities, optimized for patterns that correlate with protected characteristics, and deployed at scale across millions of underwriting decisions. As predictive models have become more prevalent in the American insurance market, regulators and researchers have increasingly recognized the potential for such models to produce unintended disparate outcomes if they are not appropriately designed, tested, monitored, and governed. Colorado has explicitly decided such outcomes are not acceptable, and the framework it has built over four years is the most serious attempt by any U.S. state to hold insurers accountable for what their algorithms actually do rather than what they are intended to do.

The Foundation: Senate Bill 21-169

Colorado's framework begins with SB 21-169, signed into law on July 6, 2021, in response to the rapid expansion of big data in insurance. Insurers had begun using nontraditional data sources, credit histories, educational attainment, purchasing habits, court records, even social media behavior, in rating and underwriting decisions. While these can correlate with actuarial risk, they also correlate closely with race, ethnicity, socioeconomic status, and other protected characteristics.

SB 21-169 prohibited insurers from using external consumer data and information sources (ECDIS), along with the algorithms and predictive models built on them, in ways that result in unfair discrimination. The first regulation under the law, Regulation 10-1-1, was adopted in 2023 and applied to life insurers.

Expanding the Reach: Auto and Health Insurance

For years, auto and health consumers sat outside the scope of these rules even as algorithms shaped their premiums and coverage. The amended Regulation 10-1-1, adopted August 20, 2025, and effective October 15, 2025, extends the full governance and risk-management framework to private passenger automobile and health benefit plan insurers.

Both must now build a comprehensive AI governance infrastructure: written policies for how ECDIS and predictive models are selected, tested, deployed, and monitored, with oversight at the board or senior-leadership level.

What Biased Algorithms Look Like in Practice

This is not an abstract concern. In stakeholder consultations during Colorado's rulemaking, it emerged that male drivers were being quoted auto premiums $58 lower than comparable female drivers, a disparity with no demonstrated relationship to driving risk.[2] In health insurance, models incorporating socioeconomic proxies may penalize low-income or minority consumers with higher premiums, reduced coverage, or claim denials. None of this requires discriminatory intent. It requires only that a model is trained on data reflecting existing inequalities, without testing whether the outputs are fair before deployment.

What This Means for Actuaries

For actuaries, Colorado's framework is a professional responsibility story. Actuaries have always been the people in insurance organizations well positioned to understand what predictive models are actually doing, where the assumptions are questionable, and where outputs may be technically defensible under one metric and genuinely unfair under another. The skills required to meet Colorado's requirements, model inventory management, bias testing methodology, documentation of testing protocols, and governance framework design, are not foreign to the actuarial toolkit. They are extensions of what sound actuarial practice already demands.

In practical terms, actuaries working with Colorado-regulated insurers may now need to be engaged in areas that may previously have sat outside their formal scope. Model governance frameworks could benefit from actuarial input on how to define and test for unfair discrimination in a statistically rigorous way. The inventory of ECDIS requires someone who understands what external data sources are actually doing inside a model, not just what the vendor claims. Bias testing requires a methodology, and methodology in a context this consequential requires actuarial judgment rather than relying on a vendor checkbox. The annual compliance reports signed by a corporate officer are only credible if the underlying work was done by people who understand what they are attesting to.

> AI governance is may be becoming a core actuarial competency.

The broader implication is that AI governance may be becoming a core actuarial competency whether the profession formally claims it or not. If actuaries do not step into this space, it will likely be filled by data scientists, compliance officers, and attorneys who may lack the combination of statistical rigor, risk-management orientation, and professional accountability that make actuarial involvement genuinely valuable. Colorado's framework is an invitation for the profession to own a problem it is uniquely qualified to solve.

Compliance Timelines and Practical Starting Points

Insurers newly covered by the regulation had to submit an interim progress report by December 1, 2025, with annual compliance reports beginning July 1, 2026. Reports must be signed by a corporate officer, and insurers falling short must submit corrective action plans.

One practical first step is a thorough inventory and gap analysis. Mapping and documenting every algorithmic model that touches underwriting, pricing, or claims is essential for building an effective governance framework. Many insurers will find this exercise reveals data and model uses senior leadership did not know existed.

> You cannot govern what you have not identified, and you cannot test for bias in a model you do not know about.

[2] Airlie Hilliard, "How Colorado Is Regulating Insurtech with SB21-169," *Holistic AI*, August 15, 2023, https://www.holisticai.com/blog/how-colorado-is-regulating-insurtech-with-sb21-169.

That is the point: you cannot govern what you have not identified, and you cannot test for bias in a model you do not know about.

Colorado as a National Model

In my opinion, Colorado's approach deserves credit for being serious. The phased rollout, stakeholder consultation, and line-specific tailoring reflect genuine regulatory craft rather than legislation written for press releases. The requirement for ongoing monitoring and annual reporting, rather than one-time certification, is especially important: bias in algorithmic systems is not solved once and put away. Models drift, inputs change, and new proxy variables emerge; a continuous accountability loop is the right response to a continuous risk.

Other states are watching. New York, California, Connecticut, and New Jersey are all considering similar frameworks. Colorado's effort positions it as the reference point others will build from and adapt. Whether you see this as an overdue correction or an administrative burden may depend on whether you have ever had to explain to a consumer why an algorithm decided they were a worse risk than their neighbor.

Syed Raza, FSA, MAAA Founder & Principal Actuarial 360 Solutions
The author offers further reading on this topic in his blog, MyActuary Weekly, *"When the Algorithm Answers for Itself: Colorado's AI Governance Framework for Insurance," May 6, 2026,* https://www.theactuaryweekly.com/p/when-the-algorithm-answers-for-itself-ai-governance-regulations-by-colorado-for-insurance.

ACTUARIAL INTELLIGENCE BULLETIN

Virtual Roundtable: AI in Actuarial Work

EDITED BY DAVE INGRAM FSA, CERA

We convened a Virtual Roundtable to hear what people are thinking about several pressing AI-related questions. With the virtual format, we were able to get participation from actuaries in 7 countries on 5 continents! This is Part 1. Part 2 coming soon.

Question 1: What is something that all actuaries should consider doing to prepare for the future with AI?

- Dakota Cooley (Cincinnati, Ohio USA)
  "The gap between people who've shipped something with these tools and people who've only read think pieces about them is enormous, and it's widening every month you wait. Stop reading about AI and actually use it on real work this week."

- Jason Reed (Boston, MA United States)
  "Develop solid background in linear algebra. Having a strong grasp of the underlying mathematics will make it much easier to understand how advanced models operate."

- Carlos Arocha (Zurich, Switzerland)
  "This means learning how AI tools work, where they fail, and how to use them

Stop reading about AI and actually use it on real work this week.

responsibly in actuarial work. The key is developing the judgment to challenge AI outputs, identify bias, protect data, explain uncertainty, and remain accountable for decisions."

- Aree Bly (Denver, USA)
  "Actuaries should be thinking broadly about various ways that AI will impact how we do our work, how AI is changing the data that we rely on for our work. While it can feel overwhelming, we can pick a few areas at a time to experiment and learn intentionally rather than trying to 'eat the elephant'."

- Mitch Stephenson (Simsbury, CT, USA)
  "Whether it is doing research, improving or automating data preparation, summarizing or writing documents, or using AI as a coding assistant, there are many practical ways actuaries already are, and should continue to, experiment with AI. Learn to use it."

Question 2: Should AI proficiency be expected of all actuaries?

- Shirly Wang (New York, USA)
  "We all will receive more and more AI-generated contents and be expected to work with AI. Knowing its limits and boundaries is a must-have skill."

- Erick Mwangi Muguchia (Nairobi, Kenya)
  "To stand behind AI-assisted work as their own, actuaries must be proficient enough to serve as expert reviewers, auditors, and certifiers of these systems. Yes, a baseline AI proficiency should be expected of all actuaries, but at varying depths depending on the role."

- Raymond Z. Lai (Kuala Lumpur, Malaysia)
  "If AI involves setting assumptions and creating models, we need to have AI proficiency to ask the right questions, validate the assumptions and models. Human involvement on judgment and decision making is required for senior actuaries who are not involved in the works."

- Ronald Poon Affat (Sao Paulo, Brazil)
  "Actuaries who cannot understand, challenge and explain AI-assisted work may remain useful specialists, but they will be less likely to move into broader leadership roles where they supervise both actuarial and non-actuarial teams. The expectation of proficiency will likely split the career velocity of the profession."

- Mitch Stephenson (Simsbury, CT, USA)
  "In the same way it is hard to imagine an actuary not being able to work with Excel, so it seems will be the case of AI in the future. Yes, the same as for most knowledge workers."

Question 3: Which areas of traditional actuarial work do you foresee being MOST significantly augmented or automated by AI?

Actuarial Work Area / Response Category	Percent of Responses
Data collection and cleansing	100%
Regulatory reporting preparation	77%
Experience analysis and trend forecasting	69%
Pricing model building and calibration	46%
Routine reserving/valuation calculations	46%
Risk Management	8%
Contract Analysis	8%
Plan Design	8%

Question 4: What do you consider the biggest benefit of AI adoption in actuarial work?

- Dakota Cooley (Cincinnati, Ohio USA)
  "We can pull from more data sources, get to insights faster, and spend more time communicating results to stakeholders in ways that drive better decisions. It takes the mundane work off our plate — data wrangling, documentation, first-pass analysis — and frees actuaries up for the judgment calls that actually require us in the loop."

- Jason Reed (Boston, MA United States)
  "The ability to do more with fewer resources, and the automation of programming tasks. This structural efficiency will let us spend less time on manual coding execution and more on validating model logic."

- Erick Mwangi Muguchia (Nairobi, Kenya)
  "The ability to re-focus actuarial talent on high-value judgement and strategic insight, rather than data wrangling and rote calculations. This elevates the profession from back-office calculators to trusted business advisors, improves job satisfaction, and can attract broader, more diverse talent pool."

- Yulia Nechay (Hadera, Israel)
  "Allowing actuaries to spend less time on repetitive work and more time on judgment and decision making. This shift will significantly optimize workflow efficiency across teams."

- Carlos Arocha (Zurich, Switzerland)
  "AI can help actuaries process large datasets, identify emerging patterns, automate routine tasks, and test more complex scenarios allowing them to spend more time interpreting results and advising decision-makers. This provides the ability to enhance actuarial judgment with faster, deeper, and more scalable analysis."

- Ronald Poon Affat (Sao Paulo, Brazil)
  "The biggest benefit is that AI can help actuaries move faster from data to insight. It frees them from routine production work so they can spend more time on judgment, strategy and communication."

AI can free actuaries from routine production work so they spend more time on interpretation, judgment, strategy, and communication.

- Michael Niemerg (Schaumburg, USA)
  "It is another means of automation — allowing us to complete more work per unit of time. This structural baseline improvement can compound over entire departments."

- Green Chen (Toronto, Canada)
  "Less building and more thinking. This basic shift fundamentally transforms the daily experience of actuarial analysts."

- Aree Bly (Denver, USA)
  "It can free up time for deeper thinking about risks that we have been too busy to focus on. This provides space to look into emerging areas that rarely get enough attention."

- Simon Hatzesberger (Munich, Germany)
  "The biggest benefit, in my view, is that AI can take over routine operational tasks, freeing up time and energy to invest in actuarial judgment where it matters most. This will change how teams allocate daily operational resources."

- Mitch Stephenson (Simsbury, CT, USA)
  "Historically, actuaries have spent a lot of time on data preparation, model building, and report generation. If done well, actuaries can automate a lot of these processes with AI and instead focus on key strategic messaging and recommendations to management."

Participants are all prior contributors to the AI Bulletin.

ACTUARIAL INTELLIGENCE BULLETIN

Prioritizing Your NIST AI RMF Effort by Practice Area

MARTIN MELCONIAN

A common mistake with NIST AI Risk Management Framework (RMF) is treating it as a uniform checklist. It is not. The four functions (Govern, Map, Measure, Manage) apply to every AI system, but the relative weights of each change dramatically with the practice area. Within each function, which subcategories carry the substantive work varies, too. A lapse model is not an underwriting model. A care management model is not a prior authorization model. A longevity projection is barely the same kind of object. The actuary's job is to know which subcategories matter most in context.

A common mistake is treating NIST AI RMF as a uniform checklist. It is not.

Prioritization matters. RMF work is substantial, and teams that scatter effort uniformly across forty-plus subcategories burn out before producing anything useful. A team that can say "for our accelerated underwriting model, the heaviest subcategories are MAP 5.1, MEASURE 2.9, and MEASURE 2.11, and here are the artifacts we produce for each" is saying something defensible. A team that can only say "we comply with NIST AI RMF" is saying nothing.

Map Failures Usually Exceed Measure Failures

In our experience, the most consequential failures concentrate in the Map function rather than the Measure function. The cleanest illustration in the literature is the 2019 Obermeyer case.[3] A widely deployed U.S. care management algorithm assigned risk scores to identify patients for high-risk programs. By AUC or Brier score on its training target (future healthcare costs), it would have passed any conventional validation.

Obermeyer, Powers, Vogeli, and Mullainathan showed the algorithm was systematically under-identifying Black/African American patients. At the 97th percentile of its risk score, Black/African American patients had 26% more chronic conditions than white patients. The composition of the auto-identified high-risk group was 17.7% Black/African American. Targeting need rather than cost would have raised that share to 46.5%. Reformulating the label reduced the measured bias by approximately 84%.

The algorithm was perfectly calibrated for cost. The bias lived entirely in the choice of label. Black/African American patients in the study population received less care for the same level of illness, producing lower observed costs, lower predicted costs, and lower risk scores. A MAP 1.1 artifact for that system should have made the intended-purpose-versus-training-target gap explicit. It did not.

This is the teaching lesson for practice-area prioritization. Where failure modes concentrate in label choice, stakeholder impact, or proxy relationships, Map dominates. Where they concentrate in predictive performance, calibration drift, or deployment stability, Measure or Manage dominates. Where they concentrate in vendor architecture or third-party data lineage, Govern dominates.

A Tour Across Four Practice Areas

Life — accelerated underwriting. A mid-size US life insurer triaging term applications using traditional features plus external consumer data sources. The fairness machinery dominates. MEASURE 2.11 and MEASURE 2.9 carry the heaviest work, with MAP 5.1 close behind for the Fundamental Rights Impact Assessment where an EU subsidiary puts the system in scope of Annex III 5(c) of the AI Act. Colorado Regulation 10-1-1, as amended in October 2025, sits directly on this deployment. GOVERN 1.6 (AI inventory) and GOVERN 6.1 (third-party ECDIS providers) need full treatment. Post-deployment, MANAGE 4.1 drift monitoring and MANAGE 4.3 incident reporting dominate.

Life — lapse and policyholder behavior modelling. Same profession, different subcategory weights. A gradient boosting lapse model used for ALM, hedging, reserving, and in-force management does not feel like an "AI" application. There is no consumer-facing decision, no decline, no individual harm. It is, however, unambiguously a material model. The legal-scope question of whether it sits inside consumer-facing AI regimes is separate from the governance-scope question, which sits squarely inside ASOP 56 model risk management regardless. The subcategories that carry the weight are MAP 1.5 (risk tolerance framed in technical-provision and hedge-effectiveness terms), MEASURE 2.5 (rolling-origin back-testing across economic regimes), and MEASURE 4.2 (quarterly experience-versus-prediction reconciliation). Fairness is a watching brief unless the model also drives retention outreach.

P&C — personal auto telematics. The practice area where fairness work is hardest — because telematics features are continuous, behavioral, and partly chosen. The question, "Where is the protected-class proxy in this feature set?" has many more candidate answers than for discrete ECDIS-based features. Late-night driving correlates with

[3] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan, "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science* 366, no. 6464 (October 25, 2019): 447–453, https://doi.org/10.1126/science.aax2342.

shift work, income, urban residence, age, and employment sector. Feature removal is rarely a clean answer. MAP 5.1 dominates through per-feature proxy analysis. MEASURE 2.7 carries unusual weight because inputs are continuous streams from devices the policyholder controls; manipulation and integrity risks are not theoretical. MEASURE 2.11 needs geographic disparate impact testing and single-feature sensitivity analysis. California's Proposition 103 regime, New York's Circular Letter No. 7 (2024), and Colorado's amended regulation all apply. The EU AI Act generally does not, because personal auto pricing is not in Annex III.

Pensions — longevity. The cleanest example of the RMF being mostly a relabeling exercise. A stochastic mortality projection model feeding trustee valuations and buy-out pricing. Fairness sensitivity is low. Privacy concerns are proportionate to standard scheme data handling rather than novel to AI. MEASURE 2.9 (explainability) dominates. The audience is a board of trustees with fiduciary responsibility, and TAS 100 Principle 4 already requires that technical work be communicated so users understand implications, materiality of judgements, and limitations. APS X2 v1.1 governs the peer review. The RMF work is largely relabeling existing artifacts as MEASURE 2.9 evidence. ASOP 27 (revised effective 2025) and ASOP 41 are the parallel U.S. standards.

Three Observations that Carry Across

Specificity is the deliverable. RMF compliance claims framed at framework level are thin. Framed at subcategory level with named artifacts, they hold up under audit.

Map failures usually exceed Measure failures in the damage they cause. Obermeyer is the canonical illustration.

The artifacts compound. A team producing a MEASURE evidence pack for underwriting in Q1, a different one for lapse in Q2, and a different one for longevity in Q3 will find the second and third substantially easier than the first. The team's vocabulary matures. Governance committee meetings become substantive rather than performative. The operational benefit is separate from any specific regulatory obligation.

> AI governance is strongest when framework requirements are connected to the actual work processes, products, and documentation used in each practice area.

The full article with the complete practice-area triage table, the cross-cutting call-outs on reinsurance, capital, and climate modelling, and expanded treatment of each vignette is in Part 3 of the NIST AI RMF for Actuaries series on the Globebyte Insights site at https://globebyte.com/insights/applying-nist-ai-rmf-across-actuarial-practice.

Martin Melconian, Founder and CEO of Globebyte

ACTUARIAL INTELLIGENCE BULLETIN

Car Damage Classification and Localization with Fine-Tuned Vision-Enabled LLMs: A GenAI Case Study

SIMON HATZESBERGER CERA, PHD AND IRIS NONNEMAN MSC AAG

Automotive claims increasingly rely on photo evidence — submitted by policyholders, repair shops, or adjusters — to support fast triage, routing, and settlement decisions. Traditional computer vision methods such as convolutional

neural networks (CNNs) perform strongly on static image classification, but they struggle to capture the contextual information that insurance applications need: localizing the damage, assessing its severity, and accounting for external factors such as lighting or weather.

To address these limitations, we employ OpenAI's GPT-4o, a vision-enabled large language model that combines image understanding with natural-language capabilities. By fine-tuning GPT-4o on a domain-specific dataset of labeled car damage images, we achieve accuracy comparable to a CNN classifier while also producing richer contextual insights — for example, distinguishing minor glass damage on a side window from a fully shattered windshield.

Approach and Techniques

Dataset and Objectives

We use a publicly available car-damage dataset from a Kaggle competition. Images are labeled into six damage classes: crack, scratch, tire flat, dent, glass shatter, and lamp broken. We draw 1,500 images and split them into training (60%), validation (20%), and test (20%) subsets. Our primary objective is supervised multiclass classification of the damage type; the secondary objective is to extract the damage location when identifiable (e.g., "windshield", "rear bumper").

Model Comparison

For the classification objective we compare three approaches, all evaluated on the same held-out test set with accuracy and weighted F1 score. The CNN baseline is trained on the training subset, tuned on the validation subset, and evaluated on the test subset. The off-the-shelf GPT-4o is applied zero-shot: each test image is passed to the model together with a strict system prompt, and Structured Outputs constrain the response to exactly one of the six classes. The fine-tuned GPT-4o is trained on input–output pairs — each input combines the system prompt with a textual representation of the image, and each output is the corresponding damage label — via OpenAI's fine-tuning platform, then evaluated identically to the off-the-shelf variant. Other multimodal LLMs, such as Google's Gemini or Meta's Llama, also support image fine-tuning and could be swapped in under the same pipeline.

In the article Advanced Applications of Generative AI in Actuarial Science: Case Studies Beyond ChatGPT, Simon Hatzesberger and Iris Nonneman demonstrate the transformative impact of Generative AI (GenAI) on actuarial science, illustrated by four implemented case studies. In the upcoming SOA AI Bulletins, we will present each case study separately. In this issue, we describe case study 3, which shows how a fine-tuned vision-enabled LLM classifies car damage types from images while also extracting contextual details such as damage location — capabilities traditional CNNs cannot provide.

Context Extraction

For the secondary objective, we adapt the system prompt and the Structured Output schema so that the fine-tuned model optionally returns the damage location alongside the class label. Other aspects of visual context that could be extracted — damage severity, weather and lighting, or vehicle make and license plate details via OCR — follow the same pattern. Because the dataset does not include location labels, predicted locations have to be verified manually.

Results

Fine-tuning (prompt engineering) GPT-4o improves its performance to a level comparable with the CNN baseline (Table 1). The fine-tuned model reaches 0.880 on both accuracy and weighted F1, outperforming the off-the-shelf GPT-4o (0.823 / 0.825) and the CNN baseline (0.837 / 0.835).

Table 1
CLASSIFICATION PERFORMANCE ON THE HELD-OUT TEST SET. HIGHER IS BETTER.

Prediction Model	Accuracy (↑)	Weighted F1 (↑)
Convolutional Neural Network	0.837	0.835
Non-fine-tuned GPT-4o	0.823	0.825
Fine-tuned GPT-4o	0.880	0.880

To demonstrate the model's contextual capabilities, we apply the fine-tuned (prompt engineering) GPT-4o to three example images (Table 2). The model correctly identifies the glass shatter on the windshield (left) and the dent on the rear bumper (right); for the tire-flat example (middle) it correctly classifies the damage but cannot determine which specific tire. Because the dataset does not include location labels, location predictions were verified manually.

Table 2
EXAMPLE PREDICTIONS FROM THE FINE-TUNED GPT-4o — DAMAGE TYPE AND (WHERE IDENTIFIABLE) DAMAGE LOCATION

			
Actual damage type	glass shatter	tire flat	dent
Predicted damage type	glass shatter	tire flat	dent
Predicted damage location	windshield	—	rear bumper

Implications for Actuarial Practice

Fine-tuning through prompt engineering a vision-enabled LLM on domain-specific data can lift classification performance above both the off-the-shelf model and a CNN baseline, and the benefits typically extend from visual to textual tasks. Beyond pure classification, vision-enabled LLMs capture contextual details and can perform optical character recognition — providing a holistic assessment that CNNs struggle to deliver. Structured Outputs ensure that predictions adhere to predefined categories, eliminating free-form inconsistencies. Beyond car damage, the approach extends to applications such as medical image analysis, fraud detection in claims and invoices, and roof damage assessment in property insurance, among others; insurance-specific benchmarks such as INS-MMbench cover many such tasks.

Vision-enabled LLMs can combine classification with richer context—such as damage location—opening new insurance applications.

LLMs are also easier to adopt than CNNs, which require expertise in model architecture and optimization; fine-tuning achieves high-quality results with comparatively little engineering. The trade-offs are non-trivial, however: fine-tuning can take considerably longer than training a CNN, and moving from prototype to production requires substantial investment in data governance, monitoring, inference cost and latency, and the privacy implications of the chosen deployment model — a third-party API, a private-cloud instance (e.g., Azure OpenAI), or an on-premise setup. The growing image-generation capabilities of LLMs also introduce fraud risks: an undamaged vehicle could be altered to show realistic damage at a

specific location and severity, potentially bypassing traditional verification without professional image-manipulation skills.

References

Full article: Hatzesberger, Simon, and Iris Nonneman. *Advanced Applications of Generative AI in Actuarial Science: Case Studies Beyond ChatGPT*. arXiv preprint arXiv:2506.18942, version 3. 2025. https://arxiv.org/abs/2506.18942.

Notebook: International Actuarial Association AI Task Force. *Actuarial AI Case Studies: Car Damage Classification and Localization*. GitHub repository, 2025. Accessed July 10, 2026. https://github.com/IAA-AITF/Actuarial-AI-Case-Studies/tree/main/case-studies/2025/car_damage_classification_and_localization.

Dataset: Nandini, Kala. *Analytics Vidhya RIPIK AI HackFest*. Kaggle dataset. Accessed July 10, 2026. https://www.kaggle.com/datasets/imnandini/analytics-vidya-ripik-ai-hackfest.

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

ACTUARIAL INTELLIGENCE BULLETIN

The No-Thinking Spell and the AI Handover

DAVE INGRAM, FSA, CERA

The world has been under a powerful spell for what is going on 4 years now. A team of powerful wizards had been working on this spell, or group of spells (they are not sure which themselves) for a long time and suddenly, the spell started to work. And it worked much, much better than they ever expected. They released the spell on the world and then the team of wizards scattered to the winds each forming a new team who repeated the work of building their no thinking spells.

They called these spells Artificial Intelligence. And they worked in the most insidious manner to stop thinking. What AI does is answer your questions. Sounds helpful, doesn't it? But even worse, it answers them in flawless language. And it will often tell you that your questions are absolutely brilliant. And that is how the spell gets you to stop thinking. Why think when AI will do it for you and at the same time will also say how smart you are for asking.

These wizards have been warning us how strong and dangerous this AI is — so strong that we must worry about an AI Takeover. But if the No Thinking spells actually work there will be no need for an AI Takeover. Instead, there will be an AI Handover. And we are moving towards that now at breakneck speed.

How Does the No Thinking Spell Work?

A good spell includes some real substantial elements, some good misdirection technique and a little bit of actual magic. The better the substance and misdirection are, the less actual magic is needed. In this cast, the substance part is a process for compressing and then very rapidly accessing a large fraction of all human experience that is available digitally. And the wizards will claim that their real magic is in the success of the compression and access of the knowledge. And like any magician, they almost never mention the misdirection parts. But in fact, it is that very misdirection that makes the whole spell work.

There are five parts to the misdirection:

1. Smooth writing delivered faster than you can read it. In perfect sentences. Perfect spelling and grammar. We humans tend to give extra credibility to smooth writing.

2. Speaking with Authority, a style that is copied from thousands of authoritative texts that it was trained upon. Why copy someone saying something tentatively?

3. Eliminate Complexity so that whatever the topic, the reader feels smart enough to understand.

4. Confidence and Plausibility, two signals that humans often use to judge credibility and AI can copy them perfectly

5. Conversational Seduction — That sycophantic praising mentioned above along with techniques like mirroring of language makes AI sound like someone you can get along with well. Why would your buddy lie to you?

These habits of communication are used by many, and they would not be much of a problem except for two things. First, that process for compressing and retrieving all human knowledge is not perfect. Sometimes AI's answer is wrong. Wrong but presented following all five aspects of misdirection. Second, with this misdirection, AI seems to be immensely useful so many are comfortable stopping their own thinking.

For the most part, the No Thinking spell is working. The response from AI reads so well, it "must" be right.

And the AI Handoff is on its way.

The Counterspells

But there is hope. There are counterspells. They do work. With these counterspells, you can be assured that you will keep thinking. Here is a working list of the counterspells:

- Careful Reading. Separate prose quality from evidence quality. Instead of asking whether something sounds convincing, ask what facts support the claim.

- Use Hard Stop Gates. Pause deliberately. Slow down. Resist the temptation to accept the first answer that appears complete. Critical thinking often requires introducing friction into a process designed to remove friction.

- Resist Closure. Many important questions cannot be resolved immediately. Maintaining uncertainty is often more intellectually honest than rushing toward a satisfying conclusion.

- Preserve Exploration. Continue looking for alternatives. Explore side paths. Investigate competing explanations rather than accepting the first coherent story.

- Human Ownership. Avoid delegating responsibility for conclusions to a machine. The final judgment remains a human responsibility.

- Questioning Techniques. Instead of treating AI responses as answers, treat them as starting points for investigation. Ask what assumptions were made. Ask what evidence was omitted. Ask what would make the conclusion wrong.

I believe those are the five best counterspells. They are personal practices and they are good practice for all.

Special Counterspells for Actuaries

To me, relying on AI in any way seems particularly inappropriate for actuaries. We hold ourselves to high standards of professionalism and responsibility for our work. We will need to slow the process down to almost human speed. The parlor trick of having the text speed past my eyes at 8 times my reading speed is not as impressive anymore anyway. The standard AI playbook tells you that if you must, you can use Human-in-the-Loop. Human-in-the-loop is not a real control process. It is exactly as successful at control as looking at a flight log after the plane lands is at controlling how a plane is flown. The human stands squarely after the execution loop.

But I also think there are three special counterspells that are available to actuaries and all other serious users of AI:

- **Human-in-the-Model (Control).** The human specifies the process and reviews the work of AI after each and every step with a requirement for real challenge.
- **Human-in-the-Flow (Collaboration).** The work is split up between human and AI according to who can do it better.
- **Augmentation**. Collaboration involves sharing the work, not AI does all the work, and you get the credit. And augmentation means treating AI as a tool, perhaps your best tool, but still a tool.

Augmentation means treating AI as a tool, perhaps your best tool, but still a tool.

Read more about these three spells in *The Actuary* magazine this fall.

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

Editorial Committee

Thank you for reading the July 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 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