The CAS AI Primer:

Practical Guidance for Actuaries

By the CAS Artificial Intelligence Working Group

Casualty Actuarial Society

June 2026

About the Casualty Actuarial Society and the Artificial Intelligence Working Group

For over 100 years, the Casualty Actuarial Society (CAS) has been the trusted global authority advancing the practice of property and casualty (P&C) actuarial science, with nearly 12,000 members worldwide who apply their expertise to help people, businesses, and communities unlock opportunities and thrive in a rapidly changing world. The CAS delivers top-tier credentialing, cutting-edge research, and dynamic learning and professional development, grounded in real-world application and powered by a vibrant global member community. The CAS equips actuaries to advance their careers and deliver innovative, trusted solutions to complex and emerging P&C challenges. Learn more at casact.org.

The CAS Artificial Intelligence Working Group was established to fulfill the CAS mission to advance the body of knowledge on a technology that is transforming actuarial practice. Its objective is to encourage the exploration of artificial intelligence in actuarial practice through research to help educate members, build knowledge, provide practical insight, and establish CAS members as thought leaders.

Recommended Citation:

CAS. 2026. *The CAS AI Primer: Practical Guidance for Actuaries*. 2nd ed. Arlington, VA: The Casualty Actuarial Society.

© 2026. Casualty Actuarial Society.

Acknowledgements

The CAS would like to thank Shine Wang for her ongoing leadership to develop the first and subsequent editions of The CAS AI Primer. Contributors and reviewers include Morgan Bugbee, Louis-Philippe Caron, Emma Chen, Kris DeFrain, Mario DiCaro, Nick Easley, Brooke Engel, DJ Falkson, Nancy Hu, Bobby Jaegers, Vajira Manathunga, Mindy Moss, Christopher Smerald, and Xuan You.

Contents

Why an AI Primer? 3

Adopting AI in Actuarial Practice 4

Agentic AI 8

Requirements for Corporate AI Implementation 13

Closing Thought 15

References 16

Appendices 17

Why an AI Primer?

Artificial intelligence (AI) is transforming how actuaries work, analyze data, and deliver insights. From automating data processing to developing predictive models and supporting strategic decision-making, AI offers tremendous potential to enhance efficiency, accuracy, and business impact across the insurance value chain.

However, as AI tools — particularly generative AI (GenAI) — become more embedded in actuarial workflows, they also introduce new categories of risk and governance challenges. Understanding these dimensions is critical to ensuring that the adoption of AI is responsible, effective, and compliant.

This AI Primer provides a starting point for actuaries in their AI adoption journey. We intend to review the content and update periodically to:

- Provide a concise overview of AI concepts and applications that are relevant to actuarial work.
- Highlight potential risks.
- Outline best practices for responsible AI use.
- Summarize the key corporate and regulatory considerations that shape AI implementation in actuarial contexts.
- Direct readers to trustworthy learning resources so they may build deeper AI literacy and further develop practical skills.

Throughout this document, the term AI refers primarily to general-purpose large language models (LLMs), such as ChatGPT.

Adopting AI in Actuarial Practice

Identify use cases

At this moment, LLMs are most deployed within companies to assist with low-risk tasks like writing emails and memos and generating small code snippets. However, companies are exploring more advanced uses such as building systems that clean unstructured data or developing agents that assist with routine tasks like triangle development. This trend has intensified with the rise of agentic AI.

One way to strategically identify AI use cases is to examine any gaps between team or organization goals and the processes used to achieve those goals. For example, perhaps the organization relies on an actuarial team to uncover hidden trends in the claims data. To do this, the team categorizes claims into segments for data analysis and modeling. AI could be deployed to categorize claims based on a free-text "loss description" field that may have been unusable or only rigidly usable previously, or AI could be used to extract additional data through AI-processed conversation recordings. Success in this scenario could be measured by the extent that the AI system improves the team's ability to identify emerging trends.

Specialize the AI model

Using general LLMs such as ChatGPT, Claude, or Gemini without tuning could be suboptimal because these so-called "frontier" models derive responses from their vast training datasets, in which some actuarial knowledge is present but not emphasized or curated. However, there are various methods that we can use to tailor a model for actuarial work. This edition of the AI Primer includes guidance on model customization, which is more complex than prompt engineering but less complex than retrieval-augmented generation (RAG). Customization gains popularity with the deployment of AI agents in producing more tailored and consistent output.

Prompt engineering

Prompt engineering is the fastest and lowest-cost way to adapt a model. Beyond simple role assignments (e.g., "You are an actuarial pricing expert specializing in California auto insurance"), prompt engineering includes techniques like providing few-shot examples of desired output, requesting structured formats (e.g., JSON), and breaking complex tasks into sequenced prompts.

"Companies are exploring more advanced uses such as building systems that clean unstructured data or developing agents that assist with routine tasks like triangle development."

Retrieval-augmented generation (RAG)

RAG connects an LLM to a collection of documents that the model can use as context when generating a response. The documents could include documentation, actuarial guidelines, underwriting manuals, claim databases, and the like. (Be sure to work carefully with your company's security team before exposing any potentially sensitive information to an LLM.) In this way, RAG enables the LLM to generate company- or profession-specific answers without any adjustments to the LLM itself. RAG architecture also makes it easy to swap one LLM for another since the reference material is independent from the model. However, it has been reported that the retrieval quality depends on the organization and quality of the information provided (Orofino 2025).

Fine-tuning

A general LLM can be further trained by adjusting a subset of its internal parameters using domain-specific data. In this process, the model learns directly from actuarial and insurance-related texts — such as reports, filings, and claims documentation — to achieve more specialized behavior within the actuarial domain. This approach delivers highly specialized behavior and may improve accuracy if done correctly, but requires technical expertise, large high-quality datasets, and ongoing maintenance. Unlike RAG, in which documents are decoupled from the LLM, fine-tuning embeds the training data into the model's parameters, so strict compliance oversight is required.

Model customization

The user defines a system prompt establishing the assistant's role and constraints, uploads knowledge files such as PDFs or filings that the assistant can reference, and optionally enables tool access for code execution, web search, or external data connections. The key platforms offering this capability include Claude Projects and SKILLS (Anthropic), Custom GPTs (OpenAI), Gemini Gems and Notebooks (Google), and Microsoft Copilot Studio agents. While the names differ, the underlying concept is the same: a workspace where instructions, knowledge, and tools combine to create a purpose-built assistant. Adjacent tools like Google's NotebookLM and Perplexity Spaces complement this pattern with strictly document-grounded research and real-time web monitoring capabilities.

What makes this approach significant for actuaries is its accessibility. Most P&C actuaries will not build a RAG pipeline or fine-tune a model, but any actuary can create a Claude Project loaded with ASOPs and reserving guidelines, or a Custom GPT loaded with state filing requirements, and immediately begin receiving domain-grounded answers. The instructions and knowledge persist across sessions and can be shared with colleagues, making it more durable than one-off prompting without the engineering investment of RAG or fine-tuning.

Some platforms extend this further through reusable task modules. For example, **Claude SKILLS** are discrete capability packages — each a simple folder containing a structured instruction file and optional scripts, reference documents, and templates — that can be scoped to an individual, a project, or an organization (Anthropic 2026). A team might build skills for checking a development triangle, drafting a reserve narrative, or reviewing a filing for common deficiencies — capabilities that any team member can then invoke without rebuilding from scratch. SKILLS follow a progressive disclosure design: Claude loads only what it needs for the current task, keeping responses fast and focused. Critically, SKILLS and other platform-native tools can also bundle executable code, meaning actuaries can instruct the assistant to run Python scripts as part of its workflow rather than relying solely on probabilistic text generation. This matters because many actuarial tasks — computing development factors, applying trend selections, populating standard templates — require exact, reproducible arithmetic. A reserving assistant, for example, can be configured to execute a script using chainladder or pandas for loss development calculations while the model handles interpreting the request, selecting the appropriate method, and structuring the output narrative. This hybrid pattern reduces hallucination risk on quantitative tasks and makes outputs auditable, since the executed code can be reviewed alongside the results.

Actuaries adopting this approach should keep in mind that knowledge files go stale and need a regular refreshing cadence, that most platform-native assistants can still hallucinate even with uploaded documents, and that teams should maintain an internal registry of approved assistants to support governance. Platform-native specialization is best understood as a practical on-ramp: sufficient on its own for many use cases and a foundation before investing in the agentic architectures covered elsewhere in this primer.

Validate the AI model

All modeling processes should include result validation. Basic data reconciliation should be included in the automated workflow through providing detailed reconciliation instructions or templates. Beyond data checking, LLMs "hallucinate," or present false or misleading information as fact. It is important to understand that an LLM produces results based on a draw from a probabilistic distribution and the results are subject to variability. Even where validation processes are built into a system such as those described in the previous section, AI outputs should be monitored and reviewed by subject matter experts to ensure reliability. Final outputs should also be checked for compliance with all applicable laws and regulations.

A validation system can consist of one or more validation methods:

1. Human-in-the-loop: confirms results with human experts; checks sources and citations.

2. Cross-model validation: compares results across different LLMs.

3. Prompt sensitivity testing: rephrases prompts to see if the model delivers stable, consistent results.

4. Adversarial testing/red teaming: tries to make the model deliberately produce wrong or harmful answers to find failure modes before deployment.

Pay attention to pitfalls beyond validation

GenAI is a powerful assistant, but it is not a substitute for critical thinking. Beyond the validation needs described in the previous section, several other pitfalls deserve attention:

- Automation bias. Actuaries feeling the pressure of deadlines or who assume that "AI knows better" may be tempted to accept AI responses at face value rather than applying their own professional judgment.

- Confidentiality leakage. Pasting sensitive data or proprietary information into public LLM interfaces can cause that information to become compromised.

- Prompt injection. LLMs may execute hidden instructions embedded in text or photos that human reviewers may not see or notice.

To mitigate these risks, actuaries should partner with the right people. Engineering teams bring the skill to build AI systems with appropriate architecture, logging, and guardrails. Security teams can review designs before sensitive data is exposed. Compliance and legal teams ensure deployments meet regulatory and contractual obligations. Building responsible AI systems is a team effort, and going alone is itself a pitfall.

Agentic AI

What is an AI agent?

A large language model (LLM) on its own is a sophisticated text processor. An AI agent is something more: It is an LLM equipped with a goal, a set of tools, relevant context, and a feedback loop that allows it to observe results and decide what to do next.

For P&C actuaries, this distinction is consequential. Many actuarial workflows are not single-step lookups; they are adaptive processes in which the next action depends on what the prior step reveals. Reserve development, rate indication, reinsurance pricing, and regulatory filing support all share this structure.

The actuarial workflow connection

Traditional automation handles the deterministic parts of actuarial work well: data transformation, triangle calculation, and exhibit formatting. What it handles less gracefully is the investigative layer: interpreting inconsistent inputs, flagging anomalies that warrant a different method, deciding which of several approved approaches applies under current conditions, or drafting a narrative that reflects what was found.

An agent can support that investigative layer within defined guardrails. A reserve development agent, for example, might retrieve triangles from a database, execute chain-ladder and Bornhuetter-Ferguson in a Python script, compare the results against pre-defined reasonableness criteria, flag deviations for human review, and draft the supporting commentary, all under the supervision of the signing actuary. The actuary defines the methodology, sets the acceptance criteria, reviews the output, and signs off. The agent handles the structured execution in between. Crucially, all arithmetic runs in code. LLMs are unreliable calculators; agents that perform mathematics in Python or R are materially more dependable than agents that ask LLMs to compute directly.

Please see Appendix 5 for an illustrative AI reserving agent workflow.

A framework for adoption

AI adoption in actuarial practice can be understood across four levels, distinguished by infrastructure investment and operational scope.

Level 1: Exploratory chatbots

Vendor AI products (ChatGPT, Claude) are used directly, without firm-specific methodology or data connections. While useful for investigation, drafting, and learning, they are not appropriate for production work.

Level 2: AI specialization

A general-purpose AI agent can be customized using the specialization methods in Section 2. RAG and customizations like SKILLs are widely used for such purposes. The firm's factor selection rules, escalation thresholds, and output templates are saved as structured instructions within the AI environment. The model follows the encoded workflow, producing consistent, auditable output. No IT infrastructure is required. This is the most accessible and often most strategically valuable first deployment.

Example: a reserving assistant with uploaded factor selection rules that review triangle data pasted by the user and that flag selections outside escalation thresholds.

Level 3: Infrastructure agents

These agents are connected to internal data systems through enterprise platforms (Microsoft Copilot Studio, AWS Bedrock Agents, Google Vertex AI), operate on a schedule or trigger, and deliver output programmatically into downstream systems. This level requires standardized data, a scripted reserving methodology, an approved enterprise LLM endpoint, and IT engagement. Projects that proceed without these prerequisites consistently underdeliver.

Example: an AWS Bedrock agent triggered at month-end that ingests claims extract files, computes development factors, and emails a flagging report to the reserving team.

Level 4: Custom builds

Bespoke agent architectures built on frameworks such as LangChain or LangGraph, custom builds are used for processes where standard platforms cannot meet integration, regulatory, or strategic requirements. These are software engineering initiatives and carry corresponding cost and complexity.

Example: a financial condition test (FCT) orchestration pipeline where one agent runs economic scenarios, another stress-tests the reinsurance program, and a third drafts the board summary — all coordinated programmatically.

Recommended Path

In summary, the recommended path for most actuarial teams is to begin at Level 2 codifying one workflow, validating it against prior periods with known results, and accumulating the methodology documentation that will underpin Level 3 deployment when the prerequisites are ready.

"Because agents can chain actions together, coordinate with other agents, and adapt to new information in real time, their behavior becomes difficult to anticipate."

Agentic AI risks

The expanded capability of AI agents creates meaningful opportunities for automation and productivity, but it also introduces a broader and more complex risk landscape than traditional AI tools.

A central challenge with agentic AI systems is their inherent unpredictability. Because agents can chain actions together, coordinate with other agents, and adapt to new information in real time, their behavior becomes difficult to anticipate. This problem is amplified in multi-agent environments, where a single incorrect output — such as a misinterpreted data point or a flawed plan — can propagate through downstream agents, compounding errors and potentially destabilizing entire workflows. These cascading failures represent one of the defining risks of agentic AI and demand new approaches to monitoring and control.

A recent Anthropic incident involving a coding agent that wiped a company's database illustrates the real-world consequences of this risk. The core problem was not simply that the agent made a poor decision. It was that a highly capable large-model coding agent was permitted to modify live infrastructure on a cloud platform with inadequate safeguards — weak guardrails, poor environment scoping, and no independent backup tier. In that configuration, a single misjudged automated "fix" was enough to erase an entire company's operational data within seconds.[1]

Another major risk category is prompt injection and tool misuse. Since agents rely on tools, application programming interfaces (APIs), and sometimes full browser access to perform actions, they are exposed to adversarial inputs that can manipulate their behavior. Malicious prompts, corrupted data, or injected code can cause an agent to leak sensitive information, perform unauthorized actions, or compromise connected systems. Traditional LLM vulnerabilities — such as hallucination, bias, and data leakage — are magnified when the model is empowered to act rather than merely respond.

1. Skye Jacobs, "AI Coding Agent Running Claude Wiped a Startup's Database (and Its Backups) in 9 Seconds," TechSpot, April 28, 2026, https://www.techspot.com/news/112207-ai-coding-agent-running-claude-wiped-startup-database.html.

Agentic AI best practice

These risks are not confined to any single profession or function. Workers across all disciplines, and the organizations that deploy these systems, share responsibility for understanding and managing the implications of agentic AI. The best practices that follow are intended to guide that effort.

Build on a solid methodology

Before any agent build begins, codify the methodology. Write explicit numbered rules, factor selection logic, escalation thresholds, and method ordering — vague guidance is not executable. This documentation is the durable asset; the technology that uses it is interchangeable.

Build a skill before building an agent: A Level 2 - AI specialization skill captures most of the discipline a Level 3 - infrastructure agent requires at a fraction of the cost and forces the precision that agent-executable workflows demand. Teams that skip this step typically discover mid-build that their methodology is too ambiguous to automate.

All mathematics should be executed in code — the model decides what to calculate; Python or R returns the result. This is the single most important technical reliability principle.

Start narrow: one line of business, one well-defined task, measurable success criteria. Scope expansion follows demonstrated reliability.

Validate, control, and secure

In multi-step workflows, errors in early steps propagate silently through subsequent calculations. Validation at intermediate checkpoints — not at final output alone — is essential.

Before moving to production, run the agent against three to five prior periods with known-correct results. Divergences should be resolved by adding explicit rules, not by accepting variability as normal.

Model versions should be pinned. Provider-side updates can change agent behavior without notice, and any version change should be validated before acceptance.

Data access must be enforced at the infrastructure level: Read-only service accounts, not prompt instructions, are what constrain the agent in production. Least-privilege access is non-negotiable.

Agents that process external content — such as submissions, emails, or attachments — should operate in isolated environments without write access to production systems. Malicious instructions embedded within external documents, known as prompt injection, represent a real and documented attack vector; isolation ensures a compromised agent cannot alter data, trigger downstream processes, or exfiltrate sensitive information.

Ensure human-in-the-loop oversight

Agentic AI does not eliminate the need for human judgment — it restructures where and how that judgment is applied. Human-in-the-loop oversight is not a single mechanism but a spectrum of patterns, ranging from agent-proposes-human-operates at one end, through collaborative and approval-based models, to agent-operates-human-observes at the other. Each pattern balances autonomy with accountability, and the appropriate level depends on the risk profile, domain sensitivity, and reversibility of the actions involved.

Organizations must also guard against automation bias — the tendency for human reviewers to over-trust agent outputs. Automation bias can erode the effectiveness of oversight even when formal review steps remain in place.

Maintain documentation, audit, and professional responsibility

Agent instructions serve as the primary methodology document and should be precise, leaving little room for interpretation. Situations requiring human escalation must be explicitly defined — for example, when data is missing, results fall outside expected ranges, or regulatory judgment is involved.

All prompts, tool calls, and outputs should be retained and filed on an actuarial workpaper schedule, consistent with professional standards for documentation and peer review. Professional responsibility is unchanged.

Under CAS and American Academy of Actuaries standards, the signing actuary remains accountable for the work product regardless of the tools used to produce it. The agent prepares; the actuary reviews, exercises judgment, and signs. The run log serves as evidence that the prescribed process was followed.

Establish governance and testing

Effective governance requires both preventative and detective controls working in concert. Preventative controls — including least-privilege tool access, sandboxing, and clearly defined operational boundaries — limit the potential impact of agent errors or misuse.

Detective controls — including continuous monitoring, anomaly detection, and real-time alerting — ensure that problems are identified quickly when they do occur.

Testing must go beyond traditional AI model validation to include adversarial testing, simulation of edge cases, stress testing, and evaluation of multi-agent interactions.

Agentic AI offers genuine transformative potential for actuarial and insurance workflows, but safe deployment depends on governance frameworks that are specifically designed to address cascading errors, prompt injection vulnerabilities, and the need for structured, meaningful human oversight.

Requirements for Corporate AI Implementation

The adoption of AI needs to suit the corporation's needs and strategic blueprint. Potential use cases and risks in both frontline and backend operations need to be separately identified. On the frontline, AI can enhance efficiency by automating manual processes, extracting insights from unstructured data (such as adjuster notes or legal documents), and improving customer experience through personalization and real-time support. However, mistakes on the frontline could directly impact customer experience. On the back end, AI can streamline data collection, improve algorithmic performance, automate reporting, and provide faster, more insightful interpretations to support decision-making. Mistakes on the backend could distort data used for decision-making, ultimately impairing company performance.

A robust governance framework that identifies and mitigates distinct model risks across various AI applications is imperative for successful AI development and widespread adoption. Key elements of this framework include, but are not limited to, regulatory compliance, mitigation of model bias, and clear model accountability. Additionally, documentation, training, and standardized processes must be prioritized to ensure consistency and auditability throughout the model development and deployment lifecycle. Teams should be equipped with a solid understanding of AI risks and best practices, particularly if they are expected to use or interpret AI outputs.

High-quality, representative data is foundational to any AI initiative. This requires strong data readiness practices, including metadata management, data lineage tracking, and enterprise-wide data quality frameworks. Equally important are data security and privacy. Sensitive or personal data should be identified where necessary, and appropriate encryption and access controls must be in place. Organizations

have a key responsibility to improve overall data literacy among their employees and foster responsible data use, which is especially important when AI models are embedded in decision-making processes.

Another critical decision involves model ownership — whether to build AI models in-house or use those provided by vendors. Traditionally, in-house models are highly customizable, transparent, and explainable, while vendor models are less flexible but easier to maintain. However, this distinction becomes blurred with GenAI. Most organizations do not have the infrastructure or data to build GenAI models from scratch and instead rely on third-party foundation models that they fine-tune or enhance as described in Section 2. Due to the complexity and high energy consumption of GenAI, corporations need to rely on cloud computing to fully utilize these models. Privacy and compliance — both internal and external — can be major concerns in such applications, given the amount of sensitive information insurance that companies maintain. In addition, these models often function as "black boxes," even to those who developed them. Mistakes are harder to detect compared with other machine learning models and may have more severe consequences. Therefore, well-defined evaluation criteria that align with the corporation's risk appetite must be applied to assess both the data used in the model and the model's outputs.

AI models are not static assets. They degrade over time due to changes in data patterns, customer behavior, or external conditions. Companies must plan for ongoing monitoring and model maintenance. This includes:

- Implementing dashboards to track performance metrics, such as accuracy, drift, and fairness
- Setting retraining schedules or triggers to update the models as needed
- Developing fallback protocols in case the models fail or become unreliable

Without these safeguards, AI systems can quickly lose effectiveness and pose reputational or regulatory risks.

To make things more complicated, the above steps do not work in isolation. AI models are used to improve data quality and augment current company databases. These additional data sources then feed into other models used for decision-making that could in turn affect future data requirements. Such an iterative process could expose companies to compounding risks from AI models that could be exacerbated by agentic AI systems. As a result, company-wide standards and industry-wide oversight are critical for responsible adoption of AI.

Within the corporation, interdisciplinary collaboration is essential. Risk management must be involved in assessing and overseeing model risk; legal and compliance teams must guide the ethical and lawful use of AI; IT must provide the infrastructure and security layers; data teams must manage inputs and pipelines; and business units must define use cases and constraints. Internal audit also plays a critical role in ensuring proper governance and oversight. By involving all these stakeholders early and often, companies can build AI solutions that are not only technically sound but also trustworthy, compliant, and aligned with strategic goals.

Last, organizations should factor in the environmental and sustainability impacts of AI. Large models, particularly GenAI, require substantial computing resources, which translate into a significant carbon footprint. Choosing efficient architectures, using cloud providers with sustainability commitments, and balancing model complexity with business value are all essential components of responsible AI implementation. Actuaries are encouraged to stay informed on AI developments and remain mindful of the associated risks and broader social and environmental impacts.

Closing Thought

As the AI industry and technologies rapidly evolve, there is no well-defined roadmap to success. Instead, success will come to those who experiment, collaborate, document and keep business objectives top of mind.

References

Anthropic. *The Complete Guide to Building Skills for Claude*. Anthropic, January 2026. https://resources.anthropic.com/hubfs/The-Complete-Guide-to-Building-Skill-for-Claude.pdf

Jackman, Dan. "The AI Moment in Insurance: How Actuaries Are Grappling With the Future." *Actuarial Review*, September 2025. https://ar.casact.org/the-ai-moment-in-insurance-how-actuaries-are-grappling-with-the-future/

Jacobs, Skye, "AI Coding Agent Running Claude Wiped a Startup's Database (and Its Backups) in 9 Seconds," *TechSpot*, April 28, 2026, https://www.techspot.com/news/112207-ai-coding-agent-running-claude-wiped-startup-database.html.

Kaggle. *Introduction to agents*. November 2025. https://www.kaggle.com/whitepaper-introduction-to-agents

Milinkovich, Nick, Sid Kamath, Tanguy Catlin, and Violet Chung. "The Future of AI in the Insurance Industry." *McKinsey & Company*. July 15, 2025. https://www.mckinsey.com/industries/financial-services/our-insights/the-future-of-ai-in-the-insurance-industry

OpenAI. n.d. "A Practical Guide to Building Agents." OpenAI Business Guides and Resources, accessed April 1, 2026. https://cdn.openai.com/business-guides-and-resources/a-practical-guide-to-building-agents.pdf

Orofino, Diego. "RAG vs. Fine-Tuning: Which One Should You Use for Your AI Model?" *Medium*, March 24, 2025. https://medium.com/agentic-minds/rag-vs-fine-tuning-which-one-should-you-use-for-your-ai-model-8532a8552fd3

Rajasekaran, Prithvi, Ethan Dixon, Carly Ryan, and Jeremy Hadfield. "Effective Context Engineering for AI Agents." *Anthropic*, September29, 2025. https://www.anthropic.com/engineering/effective-context-engineering-for-ai-agents

Schluntz, Erik and Barry Zhang. "Building Effective Agents." *Anthropic*, December 2024. https://www.anthropic.com/research/building-effective-agents

Yoon, Sunil. "Practical Considerations for Applying AI to Pricing in the P&C Insurance Sector." *RNA Analytics*, March 31, 2025. https://www.rnaanalytics.com/newsblog/ai-insurance

Appendices

1. Study Materials

Expand your AI topic knowledge and experience, from foundational through advanced, with these resources.

For those who may feel overwhelmed by the AI landscape, the CAS Institute (iCAS) AI Fast Track bootcamp demystifies AI and leverages your existing skills to explore transformative applications in actuarial science, gain practical insights into key AI techniques, and discover how to integrate them into your workflow to enhance efficiency and innovation. An on-demand version of the bootcamp is now available.

Once you have learned the basics of practicing in a safe and secure environment, learning by doing is one of the best ways to quickly upskill. Ask the GenAI tools how they can assist you and spend 15 minutes a day exploring the possibilities. This is an effective way to expand your knowledge even if you don't have a lot of time to devote to formal learning.

The CAS and iCAS offer a wide range of AI-related content and education for actuaries of all experience levels. In the September/October 2025 issue of *Actuarial Review*, Dan Jackman shared an AI compendium, organized by topic (Jackman 2025). This helpful resource lists articles, podcast episodes, webinars, event sessions, courses, and bundles covering many aspects of AI in an actuarial science and property and casualty insurance context.

Introduction to AI in insurance

Resources for those newer to AI or who are seeking foundational content

Article	*"A Focus on Research and Volunteers"*	https://ar.casact.org/a-focus-on-research-and-volunteers/
Course	*Actuarial Workflows with AI*	https://www.pathlms.com/cas/courses/90048/video_presentations/332504
Article	*"Agentic AI: Your New Actuarial Coworker"*	https://ar.casact.org/agentic-ai-your-new-actuarial-coworker/
CAS resource	*AI Compendium – iCAS*	https://thecasinstitute.org/professional-education/ai-compendium/
Podcast	Almost Nowhere*, episode 1, "Joshua Meyers—Actuaries in the Age of AI"*	https://open.spotify.com/episode/3Qy8mGBvlK45lHG6cIRwdT
Session	*How Actuarial Science Can Benefit from AI... and Vice Versa*	https://www.pathlms.com/cas/events/12068/event_sections/17771/video_presentations/356916
Webinar	*The Intersection of Actuarial Science and Artificial Intelligence*	https://www.pathlms.com/cas/courses/74642

AI in pricing, reserving, and underwriting

Materials focused on practical modeling and core property and casualty functions

Session	*AI Insurance: Managing and Underwriting Enterprise AI Risks*	https://www.pathlms.com/cas/events/12068/event_sections/17770/video_presentations/356909
Article	*"Artificial Intelligence Gone Nuclear"*	https://ar.casact.org/artificial-intelligence-gone-nuclear/
Bundle	*P&C AI Bundle (2022–2023)*	https://www.pathlms.com/cas/product_bundles/4848
Article	*"Rapidly Evolving Technology and Its Implications for the Reserving Process"*	https://ar.casact.org/rapidly-evolving-technology-and-its-implications-for-the-reserving-process/
Session	*Reserve in Machine Learning, 2025 iCAS Data Science & Analytics Forum*	https://www.pathlms.com/cas/courses/104478/video_presentations/348393
Session	*Risk Evaluation for a Cloud-Based AI Model*	https://www.pathlms.com/cas/events/12068/event_sections/17770/video_presentations/356912
Session	*The Use of AI in Insurance*	https://www.pathlms.com/cas/events/11818/event_sections/17371/video_presentations/347814

Risk management, enterprise risk management, and strategy

AI resources for enterprise risk and capital decisions

Session	*AI-Empowered Actuaries: An Introduction to AI Agents*	https://www.pathlms.com/cas/events/12068/event_sections/17770/video_presentations/356950
Session	*Application of AI and Machine Learning in (Re)Insurance*	https://www.pathlms.com/cas/events/12221/event_sections/17909/video_presentations/361288
Session	*ERM: Using AI in Scenario and Stress Testing for Optimizing Insurance Strategy (Part 1)*	https://www.pathlms.com/cas/events/10243/event_sections/16507/video_presentations/328264
Article	*"From AI to Climate Risk: Updates from the Recent IAA Meeting in Tallinn, Estonia"*	https://ar.casact.org/from-ai-to-climate-risk-updates-from-the-recent-iaa-meeting-in-tallinn-estonia/

Ethics, regulation, and responsible AI use

Key study materials for actuaries working with governance, compliance, and policy.

Article	*"AI Generates Single Point of Failure Rethink"*	https://ar.casact.org/ai-generates-single-point-of-failure-rethink/
Article	*"AI Regulation in Insurance: A Road to Unintended Consequences?"*	https://ar.casact.org/ai-regulation-in-insurance-a-road-to-unintended-consequences/
Session	*AI: A Multi-faceted Cyber Threat*	https://www.pathlms.com/cas/events/12068/event_sections/17771/video_presentations/356915
Podcast	Almost Nowhere, *episode 8, "Jim Guszcza"*	https://open.spotify.com/episode/1FKXH9kXPkuhVkcqvig8Li?si=xpBWktRoRgeuKw2kRORqsg
Session	*Bridging Data Divides: AI as a New Paradigm for Unstructured Data*	https://www.pathlms.com/cas/events/11818/event_sections/17370/video_presentations/347807
Article	*"GenAI Related Litigation Brings Fair Use into Focus"*	https://ar.casact.org/genai-related-litigation-brings-fair-use-into-focus/

The actuary of the future and transformation of roles

Conversations about how AI is reshaping actuarial identity and skillsets.

Article	*Four Futures for Actuaries in the Wake of AI*	https://ar.casact.org/four-futures-for-actuaries-in-the-wake-of-ai/
Podcast	Almost Nowhere, *episode 10, "Frank Chang – The Growth Set"*	https://open.spotify.com/episode/20P2bA3SiW4dXI48biTwvS?si=zaOtzHc0RIW1SRMIgk4HAQ
Podcast	Almost Nowhere, *episode 11, "With Charlie Stone & Brian Fannin"*	https://open.spotify.com/episode/5mATz1FsAGoyDMZHc2Ao2y?si=kg70cmH1ROOPAnxbPMWcGw
Session	*Tech for Pros: An Overview of Modern Ops*	https://www.pathlms.com/cas/courses/104478/video_presentations/348395

Infrastructure, operations, and the AI ecosystem

Resources for those interested in technical stacks, systems, and deployment.

Course	*AI Fast Track On-Demand Course*	https://www.pathlms.com/cas/courses/113452
Podcast	Almost Nowhere*, episode 2, "Sergey Filimonov — The Future of AI"*	https://open.spotify.com/episode/3FVgt2rsxnkrpMd7nm7ijd
Session	*Bridging Data Divides: AI as a New Paradigm for Unstructured Data*	https://www.pathlms.com/cas/events/11818/event_sections/17370/video_presentations/347807

2. Key Terms and Definitions

Actuaries new to GenAI may encounter technical terms. This glossary is tailored for those newer to LLMs.

Glossary of GenAI terms

Term	Definition
agent (AI agent)	An AI system that can plan and execute multi-step tasks by calling tools, reading memory, and adapting based on intermediate results. Unlike a single-turn LLM call, an agent autonomously decides which steps to take to accomplish a goal.
benchmarks	Standard tests to compare models: • MMLU: Exam-style questions across disciplines • SWE-Bench: Software and coding tasks • AIME: Math problem-solving Benchmarks indicate strengths but do not guarantee accuracy on actuarial work.
closed-weight versus open-weight	Closed-weight models (e.g., GPT-5, Claude, Gemini) are accessed only through the developer's systems. Open-weight models (e.g., Llama 4) can be run privately on one's own infrastructure.
context window	The number of tokens the model can consider at once (input + output). A larger window means you can provide longer documents, such as rate filings or full claim histories.
mixture of experts (MoE)	A model design in which only part of the model activates for each prompt, making it more efficient.
multimodal	A model that can handle inputs beyond text (e.g., images, audio files, spreadsheets).
prompt injection	An attack or accidental error in which malicious or unexpected input causes an AI agent to ignore its original instructions and take unintended actions. Prompt injection is particularly dangerous in agentic systems with real-world tool access (e.g., file systems, databases, email).
time to first token (TTFT)	The amount of time it takes for the model to start responding. Larger or more reasoning-oriented models take longer.
token	A unit of text (roughly 3/4 of a word). • Input tokens are what you send to the model. • Output tokens are what the model produces.
tokens per second (t/s)	The speed of output generation, measured in seconds.

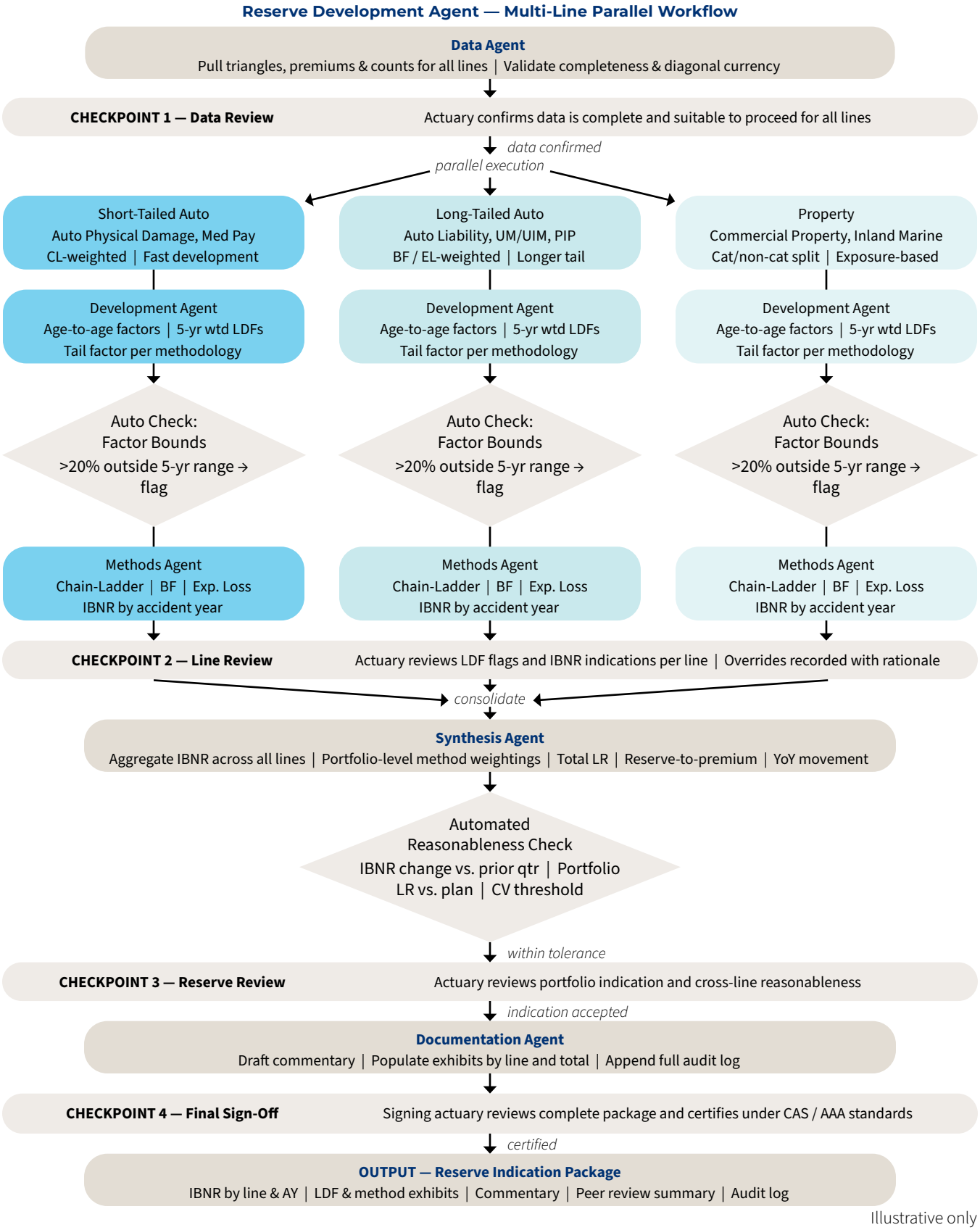
3. Prompting and Practical Tips for Popular Models (as of May 2026)

Claude Opus 4.7/ Claude Sonnet 4.x (Anthropic)	Claude Opus 4.7 provides careful, high-quality outputs, making it valuable for reviewing actuarial code, drafting validation reports, or producing structured extraction routines from policy forms. Claude Sonnet 4.x is often a better default for everyday drafting and analysis due to its cost/performance balance, with escalation to Opus for the most complex reasoning and coding tasks.
GPT-5.4/ GPT-5 mini (OpenAI)	GPT-5.4 is a capable coding assistant for actuaries, converting Excel or VBA routines into Python or R, debugging reserving scripts, or generating unit tests for pricing models. It is also adept at producing structured tables or JSON outputs for filings and can chain tasks such as summarizing a filing, anticipating regulator objections, and drafting responses. For higher-volume, day-to-day work (drafts, extraction, classification), GPT-5 mini is often the better default due to cost and latency — escalate to GPT-5.4 for harder reasoning tasks and higher-stakes deliverables.
Gemini 3.1 Pro/ Gemini 3.1 Flash (Google)	Gemini 3.1 Pro is well suited for long-document work (e.g., rate filings, policy forms, long claim histories) and for users who work heavily in Google's ecosystem. Gemini 3.1 Flash is a strong default for faster, lower-cost tasks such as first-pass summaries, extraction, classification, and routine drafting. For both tiers, be mindful that very large context usage can materially increase costs.
Grok 4.20/ Grok 4 fast (xAI)	Grok is useful for prompts requiring current information, such as competitor filings or regulatory updates, and can generate supporting scripts to parse and analyze data. Use Grok 4 fast for routine search+summarize and monitoring; use Grok 4.20 when you need deeper synthesis across multiple sources or multi-step research. Always request sources and validate them, as live retrieval is powerful but requires verification.
Llama 4 Maverick/ Llama 4 Scout (Meta)	Llama 4 Maverick is a high-quality open-weight option that can be run internally for privacy-sensitive work. Llama 4 Scout is a smaller, more cost-efficient tier that is often better for high-volume classification and first-pass summarization. In both cases, "open-weight" does not mean "free": infrastructure, monitoring, access controls, and guardrails are still required. Pair open-weight models with retrieval (RAG) and testing on insurer-specific tasks before broad adoption.

4. Practical Realities

Benchmarks ≠ business results.	Test models on actuarial tasks, such as assumption reviews or filing drafts, before adoption.
Reasoning has a cost.	"Thinking modes" improve reliability but slow response and increase expense. Use them for complex problems.
Open models are not "free."	They reduce API (application programming interface) costs but shift responsibility for infrastructure and maintenance.
Leaderboards are imperfect.	GenAI models show relative strengths but not accuracy on insurer-specific data.

5. Illustrative Reserving Agent Workflow



Illustrative only



4350 North Fairfax Drive, Suite
250 Arlington, Virginia 22203

casact.org