

# The Audit-Bearing AI Standard

---

## A framework for P&C carriers adopting AI in regulated claims work

*A perspective from Velerian, PBC · June 2026*

---

### Executive summary

A P&C carrier in 2026 has two facts to reconcile.

The first: regulators have moved from observation to active examination. The NAIC AI Model Bulletin has been adopted by more than 25 states plus DC. The NAIC AI Systems Evaluation Tool is in a 12-state pilot through September. The New York Department of Financial Services has issued specific guidance on AI in underwriting and claims. The infrastructure for examiners to inspect how carriers use AI in claims operations is being built right now.

The second: the work AI can usefully do in claims operations is real. Coverage interpretation, denial generation, fair-claims-handling review, fraud screening — these are exactly the parts of the claims function where consistent application of policy, statute, and case law produces large operational gains. The AI tools that can perform this work credibly are in production today.

The problem is that the AI tools currently marketed at claims operations were designed before the examination infrastructure existed. They optimize for speed, throughput, or single-task accuracy. Most do not produce the artifact carriers will be asked to show: a defensible record of how each decision was made.

This paper proposes a framework — the **audit-bearing AI standard** — for thinking about which claims work AI should do, and how the evidence trail must be structured to satisfy the examination regime that is forming now.

---

### The shift: from AI adoption to AI accountability

For most of the AI-in-claims discussion since 2022, the question has been adoption. Could the models do the work? How would they integrate with Guidewire or Duck Creek? What use cases would deliver ROI? Carriers' answers varied, but the question was consistent: where do we start using AI in claims?

The question has changed. As of 2026, the more pressing question is: how do we use AI in claims in a way that examiners will accept?

The NAIC AI Model Bulletin, first issued in late 2023 and now adopted in over half of US states with substantively similar guidance, sets the framing. It requires insurers to maintain governance programs around AI systems used in any part of the insurance lifecycle, including claim administration. The bulletin's scope is broad: documentation

of model purpose and design, vendor management, board-level oversight, bias and discrimination assessment, ongoing performance monitoring, and adverse outcome remediation.

What makes 2026 different is what comes next. The NAIC’s AI Systems Evaluation Tool — currently being piloted by 12 states through September 2026 — gives examiners a structured framework for inspecting carrier AI governance programs during market conduct examinations. It is the operational mechanism by which the bulletin moves from policy statement to active enforcement.

Carriers that adopted AI in claims operations before this examination infrastructure existed often did so without considering what evidence they would need to produce. Their AI tools generated decisions; they did not generate defenses of those decisions. When examiners arrive — and the pilot’s structure suggests they will arrive in the next 12-24 months — those carriers will be asked to demonstrate something their tools were not designed to demonstrate.

The shift is from AI adoption to AI accountability. The carrier-side question is no longer “should we use AI?” but “are we using AI in a way that can be examined?”

---

## The audit-bearing slice

Not every part of a claims operation is equally affected by this shift. The audit-bearing slice is the subset of claims work where:

- The output of a process affects a regulated outcome (coverage determination, claim payment, denial decision, fraud disposition)
- The reasoning behind the output is examinable by a regulator, plaintiff’s counsel, or internal compliance function
- An adverse outcome (wrongful denial, bad-faith finding, fair-claims violation) produces direct financial or reputational exposure

Concretely, the audit-bearing slice in P&C claims includes four areas.

**Coverage interpretation.** Applying policy language to specific facts to determine whether a loss is covered. State-by-state policy interpretation variation, coverage ambiguity doctrines, and judicial precedent all matter. A coverage decision that cannot be defended against the specific policy provisions invoked is a coverage decision that will be reversed under scrutiny.

**Denial generation.** Drafting denial letters that meet statutory requirements for content, timing, and tone. State Unfair Claims Settlement Practices Acts (UCSPA) prescribe specific obligations. Bad-faith doctrine adds case-law overlay that varies by jurisdiction. A denial letter that omits a required element, mischaracterizes a policy provision, or fails to acknowledge a legitimate basis for coverage is a denial letter that creates litigation exposure.

**Fair-claims-handling review.** Assessing claims handling against state UCSPA re-

quirements and internal carrier standards. The bar is not “did we technically comply.” It is “would a reasonable examiner conclude we acted in good faith.” Pattern violations — consistent delays, consistent partial payments, consistent denials of certain claim types — are the kind of evidence that drives DOI consent orders.

**Fraud screening.** Applying detection logic in a way that is both effective and fair. The NAIC AI Model Bulletin’s bias and discrimination provisions apply directly. A fraud-detection model that performs differently for different demographic groups is a fraud-detection model that will be examined.

Each of these tasks shares a common structure: a reasoned application of rules to facts, producing an outcome that must withstand external scrutiny. Each is the kind of work AI can usefully accelerate. None can be done well by an AI tool that produces output without producing the reasoning behind it.

This is what we mean by audit-bearing. The bar is not just to do the work. The bar is to produce evidence that the work was done correctly.

---

## The regulatory landscape

A practitioner-level summary of where this stands as of mid-2026.

**The NAIC AI Model Bulletin.** First adopted by NAIC in late 2023. As of early 2026, more than 25 states plus DC have adopted the bulletin or substantively similar guidance. Notable early adopters include Delaware, Hawaii, Kentucky, Maryland, Massachusetts, Nebraska, New Jersey, North Carolina, Oklahoma, Pennsylvania, Rhode Island, Vermont, and Wisconsin. The bulletin applies across the insurance lifecycle, including claim administration and fraud detection. It does not prescribe specific technology; it prescribes governance.

**The NAIC AI Systems Evaluation Tool.** Currently in a 12-state pilot from January through September 2026. The Tool gives market conduct examiners a structured assessment framework for reviewing insurer AI governance programs. The pilot is designed to produce a final version of the Tool that can be deployed by all member states. Carriers domiciled or doing business in pilot states should assume their AI claims governance can be examined at any market conduct review.

**New York DFS Insurance Circular Letter No. 7 (July 2024).** Provides explicit guidance to insurers authorized in New York on AI use in underwriting and claims. The Circular Letter requires insurers to assess their AI systems for fairness, transparency, and accuracy, and to document those assessments. NYDFS has a strong track record of using Circular Letters as the basis for subsequent examination and enforcement.

**Colorado SB 21-169.** Focused on algorithmic discrimination in insurance. While the rulemaking has been heaviest on personal auto and life underwriting, the principles apply across the insurance lifecycle. The Colorado Division of Insurance has signaled increasing scrutiny of carrier algorithmic systems.

**State Unfair Claims Settlement Practices Acts.** Adopted in substantively similar form across nearly all states. UCSPA provisions govern fair-claims handling, denial requirements, and bad-faith standards. AI systems that participate in coverage determination, denial generation, or fairness review must produce output that is consistent with UCSPA requirements — and that consistency must be demonstrable.

**Bad-faith doctrine.** Case-law overlay that varies by jurisdiction. In jurisdictions with strong bad-faith doctrine — notably California, Washington, and several others — the cost of an indefensible claims decision is not just the underlying claim payment. It can be extracontractual damages many multiples larger. AI tools that affect denial or settlement decisions in these jurisdictions cannot be deployed without an evidentiary record.

The cumulative picture: an environment where examination infrastructure is being built, where the bar for what counts as defensible AI use is being defined, and where the cost of getting it wrong is rising. The carriers that adapt early — by adopting governance-first AI rather than AI with bolted-on governance — will be positioned to operate within the regime that is coming. The carriers that do not will face examination findings, consent orders, and litigation in the regimes that have arrived.

---

## The current vendor landscape

The AI tools currently marketed to P&C carriers for claims operations were designed for an earlier moment. They fall into four categories.

**Point solutions for specific tasks.** Snapsheet for virtual claims intake and photo-based property appraisal. Tractable for damage assessment from images. Hi Marley for SMS-based claims communications. CCC Intelligent Solutions for auto claims ecosystem operations. These tools do their tasks well. None were designed to produce examiner-ready evidence of the reasoning behind each output, because that was not the bar when they were built.

**Specialized fraud platforms.** Shift Technology and Friss provide fraud scoring as a layer on top of existing claims systems. They produce useful detections. Whether the detection logic is consistent across demographic groups — the bar set by the NAIC bulletin's bias provisions — is increasingly the question. Carriers using these tools are responsible for demonstrating that the tool's outputs do not produce disparate impact.

**AI-native claims management platforms.** Five Sigma offers a multi-agent AI claims management system. The model is to replace existing claims systems with an AI-native one. The challenge is that replacing a claims system is a multi-year change-management program; most regional and specialty carriers do not have that runway. And replacing the system does not, in itself, produce the kind of governance documentation examiners are now requesting.

**Core-system embedded AI.** Guidewire ClaimCenter, Duck Creek Claims, and EIS Suite have all developed embedded AI capabilities. These tools benefit from being inside the carrier's system of record. They are also limited by being inside it: a carrier

cannot adopt the in-system AI independently of its broader system-vendor relationship, and the governance posture is whatever the core-system vendor provides.

What is missing from this landscape is a category that explicitly serves the audit-bearing slice: governed AI agents that produce examiner-ready evidence as part of every decision, that run alongside the carrier's claims system rather than replacing it, and that allow the carrier to control its own governance stack.

This is the gap. It is the gap regulators have created by changing the bar. It is the gap the next generation of claims AI tools will need to fill.

---

## A framework for adoption

For a P&C carrier evaluating AI in claims operations under the new regime, we propose a five-question framework.

**1. What slice of claims work are we evaluating?** Distinguish the audit-bearing slice — coverage interpretation, denial generation, fair-claims review, fraud screening — from supporting tasks (intake, communications, scheduling). The governance bar for the audit-bearing slice is materially higher, and the tools you select should reflect that.

**2. What evidence does this tool produce per decision?** Ask the vendor to walk you through the artifact a single agent run produces. Is there model documentation an examiner could read? Decision provenance showing what input data was used and what reasoning was applied? Bias monitoring evidence at the population level? An NAIC-aligned governance record? If the answer to any of these is “we can build that for you” rather than “here is the artifact,” the tool is not yet audit-bearing.

**3. Who controls the governance stack?** If the AI is embedded in your core-system vendor's product, your governance posture is constrained by their roadmap. If the AI runs as a separate layer you control, your governance posture is yours. For the audit-bearing slice, the latter is increasingly the requirement.

**4. How does this work with our claims-handling manual and state-specific guidelines?** Generic AI does not satisfy state-specific UCSPA requirements. The tool must be configurable against your specific policy language, your specific claims-handling manual, and the specific state guidelines that apply to your business. If the tool is built around a generic understanding of “claims,” it will fail in jurisdictions with specific requirements.

**5. What is our remediation path when something goes wrong?** Models drift. Edge cases produce unexpected outputs. Examiners ask questions that require historical reconstruction. A tool that does not allow the carrier to inspect, correct, override, and document the corrective action is a tool that fails the NAIC bulletin's remediation provisions.

These five questions are not exhaustive. They are the questions a carrier evaluating

an audit-bearing AI tool in 2026 should be able to answer before adoption — and the questions an examiner reviewing that adoption will ask.

---

## **What governance-first means in practice**

The framework above implies a posture: governance is not a layer added after the model works. Governance is the design principle that shapes how the model works in the first place.

In practice, governance-first means five things.

### **Evidence is produced at the time of decision, not reconstructed afterward.**

When an agent applies coverage analysis to a claim, the citations, the reasoning trace, the policy provisions invoked, and the governance check result are produced as part of the output. They are not generated by a separate compliance review run weeks later. They live with the decision from the moment the decision is made.

**Bias monitoring is built into the gate, not bolted onto the reporting.** Every agent run includes a bias check against the populations the carrier serves. The check fires before output is published, not after. Drift is detected at the run, not at the quarterly review.

**Source citations are mandatory.** No reasoning that cannot be cited to a specific policy provision, state statute, regulatory bulletin, or case-law precedent. Output that cannot be cited does not ship.

**Human authority is preserved.** The agent is decision support, not autonomous decision-making. Final authority on every claim decision remains with the adjuster or examiner. The agent's role is to make their work defensible — not to replace their judgment.

**Customer audit rights are explicit.** The carrier can inspect, in real time, what the agent has done on its behalf. This is not a forensic capability used only when something goes wrong. It is an operational capability used every day.

These are not abstract principles. Each one corresponds to a specific NAIC AI Model Bulletin requirement. A carrier whose AI tools satisfy these five conditions has a governance program. A carrier whose tools do not has an examination exposure.

---

## **The standard is forming now**

The conventional wisdom in 2024 was that AI regulation in insurance would take years to materialize and even longer to be enforced. The NAIC AI Model Bulletin's wide and rapid adoption, combined with the AI Systems Evaluation Tool pilot, suggests a faster timeline. The carriers that move first — by adopting tools designed for the regime that is forming — will set the standard. The carriers that wait will adopt under the standard others set.

The audit-bearing AI standard is not a single product or a single vendor. It is a posture toward AI in regulated work: produce the evidence at the time of decision, give the human final authority, make the governance check non-bypassable, and ensure the carrier controls the documentation. Tools that operate under this posture are tools that will pass examination. Tools that do not will not.

The next 18 months will determine which tools belong in which category. The choice is in front of every carrier's claims and compliance leadership now.

---

## **About Velerian**

Velerian, PBC is a Delaware Public Benefit Corporation building governed AI agents for the audit-bearing work in regulated insurance. Our first product, claimforge, is a production-ready Insurance Claims pack anchored on NAIC UCSPA, bad-faith doctrine, prompt-payment statutes, UDAP, GLBA, and fraud-detection fairness. The platform is AWS-native and runs alongside Guidewire, Duck Creek, EIS, Insurity, Majesco, or legacy claims systems — we do not replace claims systems and we do not replace adjusters.

We are selectively engaging U.S. regional and specialty P&C carriers as design partners in 2026. If your team is thinking about how to adopt AI in the audit-bearing slice of claims operations, we would value a conversation.

**Samy Hamad** — Co-founder & CEO — [samy.hamad@velerian.ai](mailto:samy.hamad@velerian.ai) — [velerian.ai](https://velerian.ai)

---

*© 2026 Velerian, PBC. This paper expresses the views of the authors and is provided for informational purposes; it does not constitute legal advice. Carriers should consult their own counsel regarding the application of regulatory requirements to their specific circumstances.*