



# Gartner Named Sayari Scout in Two AI Reports. Here's What That Signals for TPRM.

## TPRM

# Gartner Named Sayari Scout in Two AI Reports. Here's What That Signals for TPRM.

By Sayari Analyst Team · Published October 2025

Something changed this year that didn't make the compliance press - but should have.

In February 2026, Gartner released its inaugural *Market Guide for AI Evaluation and Observability Platforms* (AEOP), formally naming AI evaluation as a distinct technology discipline. The market guide cited Sayari Scout as a representative vendor. One month later, Gartner published its analysis of AI in IT vendor risk management - and named Sayari Scout again, this time for a capability that gets to the heart of why so many TPRM programs are struggling to trust AI at scale.

The numbers Gartner published alongside these recognitions frame the stakes clearly. Today, only 18% of organizations use AI evaluation tools to test the AI applications they're deploying. Gartner projects that will reach 60% by 2028. That gap - 18% to 60% in two years - is not a prediction about enthusiasm for a new product category. It's a prediction about how many organizations are going to discover, the hard way, that deploying AI without a framework for evaluating its judgment is a different kind of risk than the ones they were trying to manage.

## The Accuracy Problem That's Sitting in Most TPRM Programs Right Now

The March 2026 Gartner report on AI for IT vendor risk management is blunt about an uncomfortable reality: AI accuracy in TPRM workflows is variable - ranging from 80% to nearly 100% across vendors and use cases. For most enterprise applications, 80% accuracy is acceptable. For TPRM in a regulated industry, it may not be.

Gartner puts the stakes plainly: in banking, government, and healthcare environments, where a failure to identify a risk could result in a material business outage or regulatory action, the tolerance for automated inaccuracy is minimal. More pointedly, a negative audit finding against an automated risk process can discredit an entire risk program - not just the specific flag that

was missed. This is a different failure mode than human error, and it's why AI accuracy in TPRM isn't just a product question. It's a program integrity question.

The same report identifies a separate but related finding: 42% of significant IT vendor incidents or outages stem from unidentified risks. That's a problem AI is supposed to solve - but only if the AI itself can be trusted to identify what matters.

This is where Sayari Scout's Gartner description becomes significant. Among the vendors cited in the reliability and accuracy section of the TPRM report, Sayari Scout was described as offering "a proprietary AI/NLP inference pipeline to eliminate hallucinations." That's specific language for a specific capability - not a general claim about AI-powered features.

## What Gartner's New AI Evaluation Category Actually Means

Gartner defines AI Evaluation and Observability Platforms as tools that manage the fundamental challenge of nondeterminism in AI systems: the fact that AI applications produce outputs that can vary, degrade over time, and fail in ways that traditional software testing cannot detect.

The framing Gartner uses to explain why this is different is precise: traditional software tests check whether a system produces the right answer. AI evaluations grade whether a system is making good judgments. The former has a single correct output. The latter requires a rubric - a definition of what "good" looks like in a specific context, against which AI responses can be consistently scored.

This distinction is the entire point of the AEOP category. And it explains why domain-specificity is the primary differentiator Gartner highlights among vendors in this market. General-purpose evaluation frameworks can measure fluency, coherence, and factual accuracy against broad benchmarks. They cannot evaluate whether an AI correctly applies the OFAC 50% rule when an ownership chain is intentionally fragmented across four jurisdictions. They cannot benchmark whether an AI identifies control relationships in a VIE structure that doesn't appear in equity filings. Those require a rubric calibrated to the specific regulatory context - and built by people who understand how that context fails.

EconSecHELM, Sayari's evaluation framework, was built for exactly this problem. It benchmarks AI against the analytic standards used in intelligence community tradecraft - not against general NLP benchmarks. The questions it answers are the questions that matter in a sanctions, TPRM, or AML workflow: does the AI exercise the judgment a trained analyst would,

or does it produce plausible-sounding output that doesn't hold up under scrutiny?

## Runtime Governance Is Not Judgment Quality

One more Gartner report from this period is worth reading alongside the AEOP and TPRM analyses: the March 2026 first take on NVIDIA's NemoClaw launch, which addresses the agent governance problem from the infrastructure layer.

NemoClaw represents a genuine attempt to solve the runtime enforcement problem - keeping AI agents inside defined policy boundaries while they execute. Gartner's assessment of it is instructive: strategically important, but not enterprise-ready, and notably limited in scope. NemoClaw addresses what agents can *do* at runtime. It does not address whether agents are making *good decisions* - a capability that requires separate controls for model evaluation, tool vetting, and judgment quality assessment.

Gartner has warned that 50% of AI-agent deployment failures by 2030 will stem from weak runtime enforcement and multisystem interoperability. That's the problem NemoClaw is trying to solve. But the other half - agents that operate within their policy boundaries and still produce wrong, hallucinated, or defensibly-incorrect outputs - requires an evaluation layer. That's the problem the AEOP category addresses. And in regulated TPRM workflows, it may be the more consequential half.

The distinction matters for how compliance leaders think about their AI stack. Runtime governance and judgment evaluation are separate disciplines. A TPRM program needs both - and most organizations currently have neither.

## Three Questions to Ask Before Your Next AI Vendor Renewal

Gartner's recommendation for organizations evaluating AEOP tools applies with equal force to any AI capability being embedded in a TPRM workflow: select solutions whose differentiators directly address your domain-specific pain points, not just your general accuracy and speed requirements. The implication for compliance leaders is a different set of evaluation questions than most TPRM RFPs currently ask.

---

# The Adoption Curve Has a Deadline

Gartner's strategic planning assumption for AI evaluation platforms is worth restating in full: by 2028, 60% of software engineering teams will use AI evaluation and observability platforms to build user trust in AI applications, up from 18% today. In a two-year window, that's an enormous normalization of a discipline that barely existed as a formal category 12 months ago.

For compliance and TPRM leaders, the implication is straightforward: the organizations that establish evaluation rigor early will have defensible AI-assisted programs when regulators formalize their expectations. The organizations that don't will be retrofitting evaluation frameworks into production AI systems under audit pressure - a considerably worse situation.

The market just named the problem. The evaluation standard for AI in regulated risk workflows exists. The question is how quickly compliance programs treat AI evaluation as a program requirement rather than a vendor feature to come back to later.

*Sayari Scout was cited as a representative vendor in Gartner's Market Guide for AI Evaluation and Observability Platforms (ID: G00842253, February 2026) and named in Gartner's "Leverage AI and Analytics for IT Vendor Risk Mitigation" (ID: G00848780, March 2026). Gartner does not endorse any vendor, product, or service depicted in its research. To explore Sayari Scout and the EconSecHELM evaluation framework, request a demo or explore the resources library.*

Please visit [sayari.com](https://sayari.com) to learn more.

*This blog is for informational purposes and isn't intended to be legal advice.*