A WHITEPAPER BY

ROSÉ

The Black-Box AI Agent Problem

Why Provable Execution Is the Missing Foundation for Enterprise AI

March 31, 2026 · rosé.xyz

EXECUTIVE SUMMARY

AI agents are rapidly moving from experimentation to production. Sixty-two percent of organizations are at least experimenting with AI agents, while 23% are already scaling agentic systems in at least one business function. Gartner predicts that 40% of enterprise applications will include task-specific AI agents by the end of 2026, up from less than 5% in 2025.

Yet most enterprises still cannot provably demonstrate what these systems actually did when they screen candidates, route payments, handle claims, or trigger regulated outcomes.

Traditional logs and audit trails — designed for deterministic software and human workflows — were never built to capture the full reasoning, tool calls, context shifts, or autonomous decision paths of AI agents. This creates a dangerous gap in legal defensibility, regulatory compliance, and risk management.

ROSÉ solves the corporate black-box AI agent problem by providing a dedicated proof layer — cryptographically sealed, ordered, and independently verifiable records for consequential AI executions.

By enabling provability, ROSÉ turns accountability into an enabler of faster, safer AI adoption rather than a barrier. Organizations that can prove what their agents did will confidently scale high-value automation while reducing exposure to liability, disputes, and regulatory scrutiny.

62%	40%	21%
of organizations are experimenting with AI agents	of enterprise apps will feature task-specific agents by end of 2026	have a mature model for governing autonomous AI agents

Logs Are Not Proof

Most enterprises rely on logging, monitoring, SIEM tools, workflow history, and cloud audit trails for troubleshooting and investigation. These systems were designed for predictable, deterministic software and human-driven processes.

They fall short when applied to autonomous AI agents.

AI agents introduce non-determinism, long-horizon reasoning chains, parallel tool calls, and dynamic context shifts that traditional logging infrastructure was never architected to fully

capture. As a result, logs often become fragmented, incomplete, or insufficient to reconstruct the exact sequence of decisions, governing rules, delegated authority, or data context at the moment of action.

In the age of agentic AI, this limitation becomes dangerous.

When an AI system rejects a job candidate, routes a payment, flags a customer, modifies a contract workflow, or triggers a regulated outcome, the question is no longer simply *what do the logs say*?

> *The real question is: Can you prove what happened? — not just internally, but to regulators, courts, insurers, auditors, boards, or counterparties.*
>
> *That is a fundamentally different and much higher standard.*

The Real Black Box Is Execution

When people talk about black-box AI, they usually mean model interpretability. That matters.

In the enterprise, however, the more urgent problem is execution.

An organization may know that a candidate was rejected or a payment was initiated. What it often cannot prove cleanly is:

- Which rules governed the action
- What authority the agent had
- Whether the action stayed within constraints
- What data context mattered at that moment
- Whether the event sequence is complete and timestamps are independent and defensible
- Whether the history can later be challenged or ambiguously reconstructed

This is the difference between *observing an outcome* and *proving an execution* — the corporate black-box AI agent problem.

Why This Matters Now

AI systems are shifting from assistance to agency. Enterprises now rely on AI-mediated systems for workflows with real economic and legal consequences.

This is especially critical in:

- Hiring and workforce decisions
- Payments and treasury workflows
- Insurance and claims handling
- Customer eligibility and fraud actions
- Procurement and contract execution
- Regulated enterprise operations

Recent data underscores the urgency: nearly three-quarters of companies plan to deploy agentic AI within two years, yet only about one in five (21%) has a mature model for governing autonomous AI agents.

When AI takes action, the core question becomes: *Can the company prove what happened in a trustworthy way*? This is no longer theoretical; it is operational.

What Enterprises Actually Need

Enterprises need more than better monitoring. They need a stronger trust boundary for consequential AI actions — the ability to show:

- What action occurred
- When it occurred and in what order
- Under which governing constraints
- With what delegated authority
- In a form that is independently verifiable and resistant to tampering

> *In short, they need provability.*
>
> *Provability transforms difficult-to-reconstruct events into defensible enterprise records. It supports compliance, dispute resolution, legal defensibility, auditor and insurer confidence, internal investigations, and safer scaling of AI.*
>
> *Without it, organizations risk pushing AI into sensitive workflows faster than they can defend it.*

Where ROSÉ Fits

ROSÉ was built to solve this exact problem.

ROSÉ is a proof layer for consequential AI decisions and actions. It operates at the point where AI-mediated activity becomes economically, legally, or operationally significant.

Instead of relying solely on ordinary logs, ROSÉ generates governed, ordered, timestamped, and cryptographically sealed proof records of execution.

ROSÉ is not another dashboard or after-the-fact tool — it is infrastructure for proving execution. It binds together:

- The action
- The governing rules or constraints
- The ordering of events
- The timing of execution
- The verifiable proof artifact

The outcome is a far stronger evidentiary posture than conventional logging alone can deliver.

More Than Identity, More Than Credentials

AI agents require credentials, delegated authority, and policy boundaries. That is essential.

But identity is only half of trust. Enterprises must also know what the agent actually did.

Credentials establish	ROSÉ establishes
Who the agent represents.	*What happened when the agent acted.*

That is the missing layer.

Provability Enables AI Adoption

A common misconception is that governance and proof infrastructure slow down AI adoption. In reality, they are powerful enablers.

The more consequential AI becomes, the greater the need for systems that can be trusted, defended, and explained. Without provable execution, organizations hesitate, escalate reviews, and keep valuable automation locked behind fear and ambiguity.

Better infrastructure for accountability accelerates safe scaling. ROSÉ provides exactly that foundation, allowing enterprises to deploy agentic AI with confidence rather than caution.

The Future Belongs to Provable Systems

The next era of enterprise AI will be won not by systems that automate more, but by those that automate consequential actions and prove them.

Black-box execution is not sustainable when AI agents increasingly influence money, rights, obligations, and regulated outcomes. Proof must become part of the stack.

ROSÉ exists to make provable AI systems possible.

Ready to add provability to your agentic AI systems?

rosé.xyz

[1] McKinsey & Company, The State of AI: Global Survey 2025 (November 2025). [source]

[2] Gartner, "Gartner Predicts 40% of Enterprise Apps Will Feature Task-Specific AI Agents by 2026" (August 26, 2025). [source]

[3] Deloitte, The State of AI in the Enterprise – 2026 AI Report (January 2026). [source]