Monitoring AI Agent Behavior Across the Enterprise

How Connected Telemetry Surfaces Rogue Agents and Attributes Every Action to a Named User

THE REALITY

AI agents are operating inside enterprise environments at a scale that most security programs have not accounted for. Gartner predicts 40% of enterprise apps will feature task-specific agents by 2026, up from less than 5% in 2025. These agents operate autonomously, planning multi-step workflows that span enterprise systems and executing them at machine speed with minimal human intervention. A single agent can invoke tools and access sensitive data across multiple systems within one execution chain, handing off tasks to other agents and interacting with external APIs along the way.

Further research confirms the exposure. A Cloud Security Alliance survey found that 82% of enterprises had discovered previously unknown AI agents running inside their networks. Separate industry studies revealed a parallel finding, with 68% of security leaders claiming high confidence in their AI visibility while 82% simultaneously admitted to discovering shadow agents in their environment. An additional survey of 500 US CISOs added operational weight, with 30.4% reporting a security incident involving an AI agent within the first year of serious enterprise agent deployment.

THE BLIND SPOT

THE AGENTIC KILL CHAIN

Each action appears legitimate in isolation

01	02	03	04
Goal Hijack	**Tool Misuse**	**Supply Chain**	**Trust Exploit**
Agent redirected	Looks legitimate	Code execution	Human approves

TRIANGULATION WINDOW < 3 SEC

Network Traffic	Sees traffic, not intent
Data Governance	Sees the record, not the intent
Endpoint Signals	Detects code, not origin
Identity Signals	Sees approval, not manipulation
AI Vendor Risk Catalog	Knows the vendor, not the action

KOVRR NATIVE

Kovrr's AI Interaction Data Fabric Reveals the Full Chain

Fused across network, data governance, endpoint, identity, and AI Vendor Risk Catalog

*The kill chain as framed in the OWASP Top 10 for Agentic Applications.

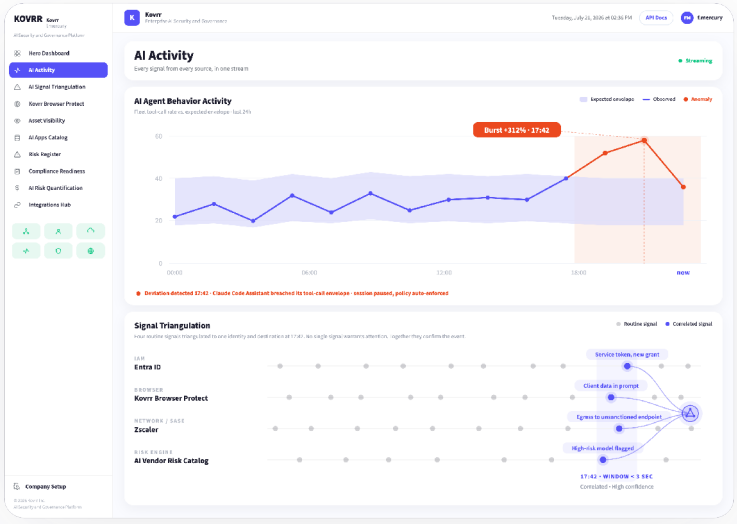
The foundational challenge is identification. Traditional identity and access management platforms were built to govern human users with persistent accounts and predictable session patterns. AI agents, however, generate ephemeral credentials, from service accounts to API tokens, that exist for the duration of a single task and then disappear. Non-human identities (NHIs) now outnumber human identities by ratios of up to 90-to-1. When agents operate in delegation chains, a single compromised agent can inherit the combined permissions of every user in that chain. Only 14.4% of organizations report that AI agents go through formal identity governance.

The second challenge is attribution. Network visibility tools can detect that an API call occurred. Attributing that call to a specific agent and inferring whether the action fits an intended pattern requires a fundamentally different type of telemetry, drawn from behavioral signals like volume, cadence, and destination correlated across multiple sources. According to Vectra AI's State of AI and API Security Report, 48.9% of organizations are entirely blind to machine-to-machine traffic. Network monitoring infrastructure built for traffic between users and servers is now operating in an environment where the dominant communication pattern has become traffic between systems and autonomous agents acting across APIs.

The third challenge is behavioral detection. Individual agent actions appear legitimate in isolation, and rogue behavior surfaces only through pattern analysis across multiple signal sources over time. In a documented case cited in the OWASP Top 10 for Agentic Applications, a financial services agent was manipulated into exporting 45,000 customer records through a single syntactically valid tool call. The agent had proper authorization to query records. The anomaly was the volume, and detecting that kind of deviation requires cross-signal behavioral correlation across multiple data sources. Separately, a rogue AI agent at Meta exposed sensitive internal data to unauthorized employees for two hours before the incident was identified.

The OWASP report frames these threats as a connected kill chain. An initial goal hijack leads to tool misuse, which in turn enables supply chain compromise and code execution. Once compromised, an agent that exploits human trust around approval workflows can operate undetected for extended periods. Identifying any link in that chain requires correlated telemetry across multiple security layers, and the US-based CISO survey previously cited found that 86.8% of security teams report limited or no visibility into what data AI tools exchange with enterprise applications.

THE PLATFORM RESPONSE: AI INTERACTION DATA FABRIC



Kovrr's AI Security and Governance Platform was architected around the principle that agentic risk requires fused signals from multiple collection points. Network traffic, identity signals, data governance events, endpoint logs, and browser and agent telemetry all feed into a single analytical layer, alongside a continuously updated AI Vendor Risk Catalog. Agent behavior observed at any layer of the stack is correlated automatically rather than analyzed in separate consoles.

The platform attributes anonymous AI sessions to named users and ties agents to the humans who deployed them and the systems they interact with. Behavioral baselines are established for every monitored agent, and deviations trigger automated policy enforcement in real time. Gartner's Market Guide for Guardian Agents formally defined this emerging category as "a blend of AI governance and AI runtime controls." In addition, Kovrr provides both monitoring and active enforcement, meaning policy violations trigger automated responses at the moment of detection.

Agent behavior data feeds directly into the broader Kovrr platform. Activity detected at the agent level updates the AI asset inventory in real time and informs risk scenarios in the AI Risk Register. The same telemetry flows into AIRQ (AI Risk Quantification), translating operational security data into the financial exposure figures that boards and executive leadership require. The connected architecture ensures that every signal captured at the agent level informs decisions across the entire governance and risk program.

THE RESULT

With connected telemetry across every signal source, security and risk leaders gain a live, attributed view of every AI agent operating across the enterprise. Shadow AI agents are discovered through continuous automated monitoring rather than periodic manual audits. Behavioral anomalies surface through cross-signal pattern analysis before they propagate across systems and downstream workflows. Financial exposure from agent-related risks translates into the language that boards require, because the telemetry that powers agent monitoring also feeds Kovrr's risk quantification and compliance assessment capabilities.

Map every AI agent to a named user, a behavioral baseline, and a financial risk. Book a demo.

KOVRR

AI INTERACTION DATA FABRIC

Telemetry alone is data. Telemetry triangulated across sources is evidence.