

GUIDE

Know what your Agents can reach. trust what they do.

AI agents are already inside your business - reading files, calling APIs, moving data - often faster than anyone can track. Some are sanctioned. Many aren't. And the ones with the widest access are frequently the ones nobody's looked at closely.



Cyera Agent Guardian secures AI agents by starting with the data they can reach and the identity they use. It finds every agent your organization is running - sanctioned or shadow - maps exactly what sensitive data each one can touch, and steps in, in real time, the moment one acts outside its lane. Security teams stop hand-auditing permissions. CISOs get a defensible number for the board. Governance and risk teams get an audit trail that actually holds up to a regulator.

Agent Guardian discovers, governs, protects, and validates - four moves, working together. Where that matters most depends on where your risk actually lives. Here are the five places we see it first.

01

Shadow AI Discovery & Governance: Find the AI tools and agents nobody approved - before they find your data

Personal ChatGPT, Claude, Perplexity, DeepSeek, local Ollama installs - and the agents employees are quietly building on top of them, from custom GPTs to low-code agent frameworks. Dozens of unapproved AI tools and agents, and company data leaving through every one of them.

- Discovers thousands of AI apps in the browser and classifies each as sanctioned or not
- Real-time blocking on the AI tools and agents used most, site-level blocking on the rest
- Inventories shadow agents alongside the tools themselves - not just the chat interface, but what's acting on top of it
- Extends to locally installed AI apps, models and agents on the endpoint, and to the network layer

Why Cyera's different: one comprehensive AI inventory and policy plane across browser, endpoint, and network - not disparate security tools that don't talk to each other.

In one engagement: ~40 shadow AI tools traced back through credit-card statements alone - tools that had never surfaced any other way.

02

Protect Developer AI Tools - Let developers use Claude Code, Copilot Studio and Cursor without exposing your source code to the internet

AI coding assistants can read source code, secrets, and credentials - and send them to external models. Security needs a way that doesn't slow engineering down.

- Discovers local AI tools on developer machines, plus local MCP servers and file-access patterns
- Blocks source code, API keys, and credentials from reaching public models - without blocking the tool
- Extends the same control to browser-based coding assistants

Why Cyera's different: Cyera identifies the actual code being touched, so it can block the secret and still allow the general question.

What this looks like in practice: Engineers shipping ~50% faster, support tickets down 25-35%, with zero code or customer data reaching public AI.



03

Detect and Remediate Over-Permissioned AI Agents - Close the data leakage path before an agent finds it

AI agents and service accounts accumulate access over time - HR data, financials, customer records - until nobody can say what any of them can actually reach, or what happens when that access gets misused: personal use of a company agent, or someone pulling a colleague's HR file - a violation that's already happened by the time anyone notices.

- Maps every AI agent and service principal to the sensitive data it can access
- Surfaces agents whose permissions exceed their actual job - the exact gap that turns into data leakage
- Lets data owners fix it directly - reduce access, revoke permissions, done
- Quarantines an agent instantly with a kill switch - stopped the moment it's flagged, not on the next review cycle

Why Cyera's different: the signal is permissions × actual data sensitivity, not access scope on its own.

In one rollout: 270,000 sensitive records pulled out of org-wide sharing - a pilot went live in under two weeks instead of a quarter-long cleanup.

04

Give auditors the trail they're actually asking for - Secure AI in Regulated Environments

Healthcare, retail, financial services and others need to prove, not assert, that AI tools are governed and that sensitive data isn't leaking through them.

- Tracks what every agent actually does with PHI, PII, and PCI - each prompt, each response, each file touched - mapped against HIPAA, PCI, and SOC 2
- Every AI asset is mapped to ISO/IEC 42001, NIST AI RMF, the EU AI Act, and the OWASP Top 10 for LLM Applications frameworks, so an agent finding influences the framework score.
- Blocks or redacts regulated data the moment it enters or leaves an agent interaction, in real time
- Timestamps every agent action down to the second, so you can show a regulator exactly when an action happened
- Keeps an immutable, agent-by-agent record of every action, policy decision, and remediation

Why Cyera's different: one continuous chain of custody - classification to enforcement to audit trail - instead of a story stitched together across five tools.

What this looks like in practice: every AI session correlated to the exact data used, across internal and external identities.



Govern every agent, wherever it runs - One inventory across Copilot, Bedrock, Claude and beyond

AI agents don't live in one place - they're built into Copilot Studio, spun up on AWS Bedrock, running on Azure AI Foundry, sometimes all three at once. Security usually ends up with a different console, and a different blind spot, for each one.



Microsoft 365 / Copilot Studio

- Discovers every Copilot Studio agent, its connected knowledge sources, and the sensitive data it can reach
- Flags overly broad SharePoint/OneDrive access, missing guardrails, or unapproved knowledge bases
- Monitors prompts and responses in real time, with inline blocking in the browser
- Quarantines an agent outright - a kill switch to block it entirely the moment it goes out of policy, not just the one risky call



AWS Bedrock

- Discovers every Bedrock agent, knowledge base, vector DB, execution role, and tool connection
- Flags the combinations that matter most: sensitive data access paired with the ability to talk externally
- Maps the full dependency chain from agent to data, visually



Claude

- Discovers every Claude deployment - claude.ai, Cowork, and Claude Code - along with every connected MCP server and custom skill, sanctioned or not
- Flags unsanctioned MCP connections, untrusted integrations, and skills with hardcoded secrets or unvetted actions before they run
- Monitors prompts, responses, and tool calls in real time, blocking out-of-policy actions before they execute



Across clouds, together

- One inventory spanning every platform, one policy engine enforced the right way on each
- The full dependency chain mapped regardless of which cloud an agent lives in
- Consistent posture checks - missing guardrails, over-permissioned roles, exposed data - no matter where you look
- Tracks every action an agent takes down to the sub-activity level, including the tokens each one costs - one place to see what agents are doing and what they're costing you

Why Cyera's different: risk scoring is grounded in what's actually behind each agent, not just how it's configured - and it's the same view whether you're looking at one platform or all three.

What this looks like in practice: 400+ Copilot Studio agents surfaced in a single deployment that nobody knew existed.



Why teams choose Cyera

It starts from the data, not just the traffic or the logs - so a finding carries the actual sensitivity behind it. Posture and runtime live in one framework instead of separate tools. And identity governance is grounded in real data sensitivity, not just role or policy - something no identity-only or data-only vendor can claim.

Not sure which of these is you?

If your agents only generate responses today, you're likely closest to the data-leakage and posture use cases. If they're already taking actions - writing, calling APIs, triggering workflows - the runtime, permissions, and production-resilience use cases are where to start. Either way, Cyera can walk through what a proof-of-value against your specific scenario looks like.

Learn more at www.cyera.com, or follow Cyera on [LinkedIn](#).



For more information
schedule a demo today

Book a demo

