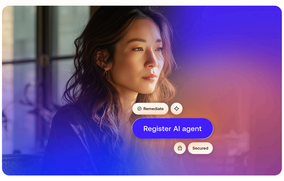
The blueprint for the Secure Agentic Enterprise

Three questions every security and IT leader must answer before agents scale beyond control



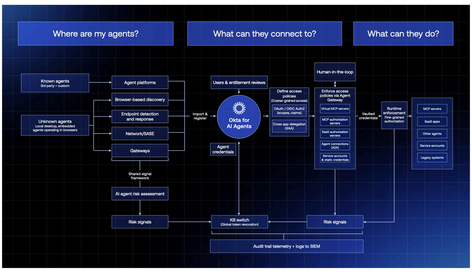
The identity gap at the center of AI security

Software used to do what you told it. Now it decides, acts, and connects on its own. AI agents are already helping employees, serving customers, and operating across supply chains—and they're scaling faster than the security around them. Over the past decade, organizations strengthened identity security for humans with vaulting, least privilege, and continuous authentication. But the rapid rise of AI agents is creating a new identity gap. Anyone can spin up an agent, agents can spawn more agents, and each one connects across apps, APIs, SaaS tools, and data systems. The result is thousands of new entities with privileged access operating at machine speed—often outside existing security controls.

That's why agents must be treated as a first-class identity. A secure agentic enterprise starts by establishing clear accountability and visibility before agents scale out of control. Organizations must be able to answer three questions:

1. Where are my agents?
2. What can they connect to?
3. What can they do?

These questions define the operational blueprint for securing AI agents. Answering them isn't just about taking inventory—it requires the right systems, identity controls, and governance model to safely operate a secure agentic enterprise.



Question 1: Where are my agents?

Visibility is the number one concern we hear from customers.

Ask your team how many agents are in your environment. Most can't give you an accurate number. The agents that employees spin up in a browser or that run quietly on desktops are often unknown and uncontrolled.

You need the ability to discover agents no matter where they were built or deployed.

- **Agentic platform integrations:** Register agents from major third-party platforms and your own custom-built agents into your identity provider. If you can't see them at creation, you can't govern them.
- **Browser-based detection:** Discover shadow agents operating through browsers and extensions that sit outside your identity provider. These are the agents employees spin up without asking permission.
- **Endpoint detection:** Identify agents running on managed devices. This should integrate with whatever you use for mobile device management or endpoint security.
- **Network detection:** Spot unauthorized agent-to-agent and agent-to-resource traffic at the network layer. Agents talk to each other and to services, and you need to see those connections.
- **Gateway detection:** Identify and govern unregistered AI agents and OAuth clients interacting with your API, MCP, and agent gateways. If agents are calling your APIs, they should be authenticated and logged.
- **AI agent risk assessment:** Continuously monitor and identify misconfigurations that leave AI agents vulnerable to exploitation. By analyzing the security posture of every agent identity, proactively surface risks.

No matter what stack you're using, you need to ingest signals from all these sources. Your discovery layer has to work across platforms, tools, and teams. Fragmented visibility is no visibility.

Question 2: What can they connect to?

Once you can see an agent, you need to map every resource it can reach and enforce access policies. Without centralized control over connection paths, a single compromised agent chains access across your environment at machine speed.

The blast radius of a compromised agent is defined by its connections.

- **MCP servers and resources:** Model Context Protocol servers give agents access to tools and data sources—both internal (apps, APIs, databases, intellectual property) and external (third-party MCPs like Slack, GitHub, and Notion). Your security perimeter now extends to every resource your agent can reach.
- **SaaS applications:** Agents connect to the same SaaS tools your employees use daily. The difference is that agents work much faster and access more data. A compromised agent can exfiltrate data or make changes across every connected SaaS app faster than a human ever could.
- **Agent-to-agent connections:** Secure the handshake and authorization between autonomous entities. One agent calling another agent is where lateral movement starts. Both sides need to verify identity before exchanging data or delegating tasks.
- **Service accounts:** Eliminate the sprawl of static, long-lived credentials. These are the skeleton keys agents inherit from legacy machine-to-machine patterns. Every static credential is a persistent backdoor waiting to be exploited.
- **Vaulted credentials:** Protect and rotate secrets automatically. An unrotated token is an open door for attackers. Credentials should be vaulted, issued dynamically, and rotated frequently. No agent should operate with a credential that outlives its task.

All of these connections need to be logged to your SIEM. You can't secure what you can't see. Every agent connection, including what it accessed, when, and with what credentials, should flow into your security operations center for monitoring and investigation.

Question 3: What can they do?

Knowing where agents are and what they can connect to isn't enough if you can't control and cut off what they actually do. When an agent starts exfiltrating data or spawning unauthorized processes, you need to respond fast.

- **Kill switch:** If an agent deviates from its intended mission, accesses sensitive data unexpectedly, or when a threat is detected, you need to be able to revoke access instantly across every system to contain risk.
- **Runtime enforcement:** Authorize agents based on what they're trying to accomplish in real time. Evaluate context, sequence, and volume. A query for 10 customer records looks different from a query for 10,000. Detect prompt injection attacks and enforce policies at the tool level before actions execute.
- **Agent lifecycle management:** Agent permissions that make sense on day one rarely make sense on day ninety. Continuously review access to enforce least privilege, automate certifications, and revoke access immediately when agents are decommissioned or employees leave.
- **Human-in-the-loop approvals:** Require human approval for sensitive or potentially risky agent actions. Prevent destructive operations, bulk data access, or agent privilege escalation.
- **Audit logs and telemetry:** Every agent action needs to be logged and sent to your SIEM. Every tool call, every authorization decision, every access attempt. Runtime enforcement and kill switches only work if you have complete visibility.

The Blueprint Is the Baseline

The three questions—where are my agents, what can they connect to, what can they do—aren't aspirational. They're the minimum standard you need for operating AI agents in production.

Organizations that can't answer them are running blind. And when the board asks, when the auditor asks, when the breach happens and the regulator asks—"I don't know" isn't an answer.

The fast movers already see this. Leading enterprises aren't waiting for the first incident to force their hand. They're treating agents as first-class identities now. They're building discovery, enforcement, and governance into their agent deployments from day one. They're answering the three questions before their agents scale beyond control.

The blueprint isn't a one-time audit. Answering the three questions is an ongoing operating discipline as your agent population grows from dozens to thousands. You spent a decade building identity security for humans. Don't let agents undo it.

Okta built this blueprint based on continuous work with leading enterprises securing AI agents at scale. Learn how the Okta Platform implements it at okta.com/ai-agents.

Disclaimers
Any mention in this white paper of solutions, features, functionalities, certifications, authorizations, or attestations that are not currently generally available or have not yet been obtained may not be delivered or obtained on time or at all. We assume no obligation to deliver on such items and you should not rely on them to make your purchase decisions.
These materials are for general informational purposes only and do not constitute legal, privacy, security, compliance, or business advice. The content may not reflect the most current security, legal and/or privacy developments. You are solely responsible for obtaining advice from your own legal and/or professional advisor and should not rely on these materials.
Okta makes no representations or warranties regarding this content and is not liable for any loss or damages resulting from your implementation of these recommendations. Information on Okta's contractual assurances to its customers may be found at okta.com/agreements.

Some images on this page were generated using the AI tool Midjourney and are used for illustrative purposes.