okta

A field guide to securing AI agents

Why identity-first security reduces hidden blind spots and helps you regain control over your automated workforce

AI agents drive better business. They also invite higher risk.

The enterprise identity landscape has shifted. While organizations have long managed non-human identities like service accounts and automated workflows, a new, more complex entity has entered the environment: autonomous AI agents. These agents have quickly become the silent engines of modern innovation, driving productivity at a scale that human teams alone cannot match.

However, this rapid growth has outpaced traditional security. Unlike standard software or basic automations, AI agents operate with a level of decision-making power and data access that creates a high-stakes attack surface. They do not just follow static scripts; they interact with sensitive data, move between systems, and execute tasks faster than legacy oversight can track.

To thrive in an agentic future, organizations can no longer treat these entities as background processes. A unified, identity-first strategy allows organizations to reduce the gap between rapid automation and comprehensive visibility.

This guide explores how to bring AI agents under the same rigorous control as your human workforce, turning shadow agency into a managed, secure advantage.

Over 80%

of all breaches now involve compromised identities, with autonomous agents representing the fastest-growing segment of that attack surface.

Source: Okta's AI at Work Report 2025: Securing the AI-powered workforce

60%

of organizations have already experienced a successful cyberattack resulting from compromised non-human identities.

Source: Okta's AI at Work Report 2025: Securing the AI-powered workforce

okta

A new generation of risk

As organizations lean into agentic workforces, the sheer volume of autonomous actors is reaching a breaking point. While standard software provides the connectivity for digital business, the move toward autonomous AI agents introduces a paradox: the more we empower them to drive productivity, the larger the security gap becomes.

Because AI agents behave more like a digital workforce than traditional software, they need a very different security approach. They often operate with broad permissions to move data between systems like AWS, Salesforce, and Jira at machine speed and without constant human oversight.

This autonomy creates a high-stakes attack surface: if an agent's identity is compromised, an adversary can leverage its existing trust to access sensitive data. Moreover, because these agents make autonomous decisions that cannot always be planned for, they risk exposing data or triggering actions that fall outside of intended business logic.

This unpredictable behavior introduces three primary vulnerabilities that demand an identity-first response:

Prompt Injection and Intent Manipulation

Attackers may manipulate prompts or inject malicious content to trick an agent. By subverting natural language inputs, an adversary can cause an agent to leak sensitive data or perform unauthorized transactions that appear legitimate to the system.

Shadow Agency and Permission Escalation

Agents are often given broad, static permissions to ensure they don't "break" during complex tasks. Without identity-first governance, these over-privileged actors become high-value targets for lateral movement and undetected network navigation.

Data Overexposure

AI agents often lack the situational context to know what not to share. Without a rigorous, identity-aware authorization framework, an agent can inadvertently synthesize and expose sensitive data to unauthorized users or external endpoints.

91%

of organizations are already using AI agents—not just investigating them, but actively deploying them.

Source: Okta's AI at Work Report 2025: Securing the AI-powered workforce

81%

of organizations do not have a well-developed strategy or roadmap for managing non-human identities (NHIs).

Source: Okta's AI at Work Report 2025: Securing the AI-powered workforce

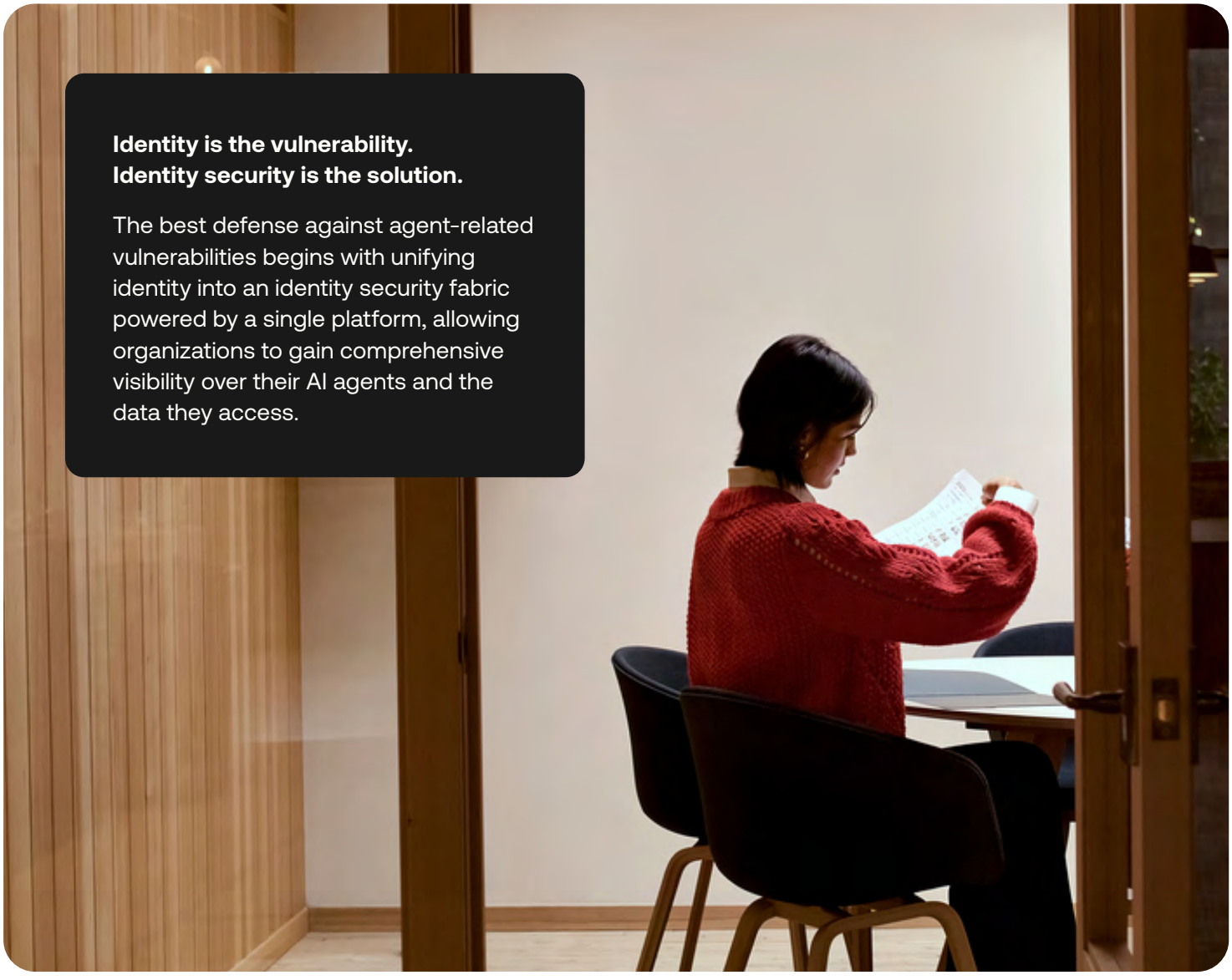
okta

The unique risk of AI agents

We have moved beyond single-use tools into the era of multi-agent orchestration, where a lead agent can autonomously spawn dozens of specialized "sub-agents" to execute complex workflows across multiple systems.

This self-multiplying autonomy creates a state of agentic sprawl, a hidden infrastructure of identities that often carry over-privileged, static permissions and operate without traditional human oversight. Because these agents move at machine speed and frequently lack the situational context to recognize unauthorized access, continuous, identity-first authorization becomes essential for maintaining effective control and security.

Vulnerabilities slip through the cracks when identity functions are scattered across different systems. If a lead agent is compromised, its ability to spawn sub-agents becomes a "golden goose" for bad actors. This allows an adversary to escalate privileges and move laterally across an ecosystem using the agent's inherited trust. For the agentic enterprise, authorization is no longer a one-time gate; it is a vital, continuous requirement for safety.



Identity is the vulnerability.
Identity security is the solution.

The best defense against agent-related vulnerabilities begins with unifying identity into an identity security fabric powered by a single platform, allowing organizations to gain comprehensive visibility over their AI agents and the data they access.

okta

Okta makes it possible

The Okta platform provides a unified approach to managing the AI agent lifecycle. By bringing agents into an identity security fabric, Okta reduces hidden blind spots and provides a comprehensive view of where agents exist, what they can connect to, and exactly what they are authorized to do.

The blueprint for a secure agentic enterprise

Okta for AI Agents provides the identity foundation required to govern AI agents at scale. Our blueprint focuses on three critical layers of control.

1. Where are my agents?

Most organizations lack a complete inventory of their agentic workforce. Agents often spin up in browsers, on endpoints, and inside isolated agent platforms without security oversight.

- **Discovery and risk detection:** Automatically find agents across your environment through integrations with agent platforms and browser-based detection.
- **Assessment:** Instantly identify "shadow agents" and prioritize remediation for those with high-risk access to sensitive data.

2. What can they connect to?

Once your agents are discovered, you must manage their access to internal and external resources. Every connection must be routed through the identity layer to prevent unauthorized lateral movement.

- **Secure interoperability:** Govern connections to SaaS applications and MCP servers.
- **Vaulted credentials:** Reduce the risk of hard-coded secrets by using vaulted, ephemeral tokens that help ensure agents do not bypass security protocols.

3. What can they do?

Securing an agent requires constant runtime enforcement. You need the ability to govern an agent from its creation to its retirement—and the power to stop it if it behaves unexpectedly.

- **Runtime enforcement:** Apply fine-grained authorization to help ensure agents only take actions they are explicitly permitted to perform.
- **Human-in-the-loop:** Use Async Auth to require human sign-off for high-consequence tasks, such as large financial transfers or data deletions.
- **The kill switch:** Manually deactivate the agent instantly with a kill switch to prevent new tokens requests and future authorizations when an agent behaves unexpectedly.

okta

Okta + Amazon Bedrock AgentCore

The key to building, securing, and scaling AI agents

AI agents are transforming how organizations operate. This environment of rapid change is a double-edged sword: For every new opportunity agents create, they also introduce new risks.

Together, Amazon Bedrock AgentCore and Okta help developers mitigate these risks while accelerating innovation. Our joint solution helps secure access and interoperability, upholds Zero Trust principles, and simplifies the management of agentic environments.

AgentCore helps you stand up purpose-built agents

- **Build and manage agents** that connect to your existing tools and services
- **Deploy with minimal or custom code**, supporting complex, multi-step agent tasks
- **Monitor agent performance** to gain insights into organizational value
- **Deliver intelligent, personalized experiences** that your customers and employees will love

Okta for AI Agents helps you securely manage your agents secures them

- **Authenticate agents** to protect sensitive customer and organizational data
- **Enforce security standards** with automated token management
- **Enable human-in-the-loop approvals** for critical or sensitive actions
- **Help ensure agents can only access what they need** with fine-grained permissions
- **Bring agents into certification workflows** to enforce least privilege and enable automated access reviews and human ownership for full lifecycle control

okta

Put your AI agents to work – securely

In record time, AI has gone from a hypothetical future advantage to a present necessity. Making full use of modern tools like AI agents is table stakes for organizations to remain competitive. However, the effective use of these tools hinges entirely on your organization's ability to secure them.

Okta brings the identity security fabric to life, securing AI-powered workflows through a unified identity platform. This approach reduces risky fragmentation and strengthens the foundation of your security ecosystem.

That's why AWS and Okta are working together to provide a secure foundation for the next generation of intelligent automation. Whether you are deploying AI agents in AWS or leveraging Amazon Q across teams, Okta can help to ensure those identities remain secure, governed, and auditable.

About Okta
Okta, Inc. is The World's Identity Company™. We secure AI, machine, and human identity so everyone is free to safely use any technology. Our customer and workforce solutions empower businesses and developers to protect their AI agents, users, employees, and partners while driving security, efficiencies, and innovation. Learn why the world's leading brands trust Okta for authentication, authorization, and more at okta.com.