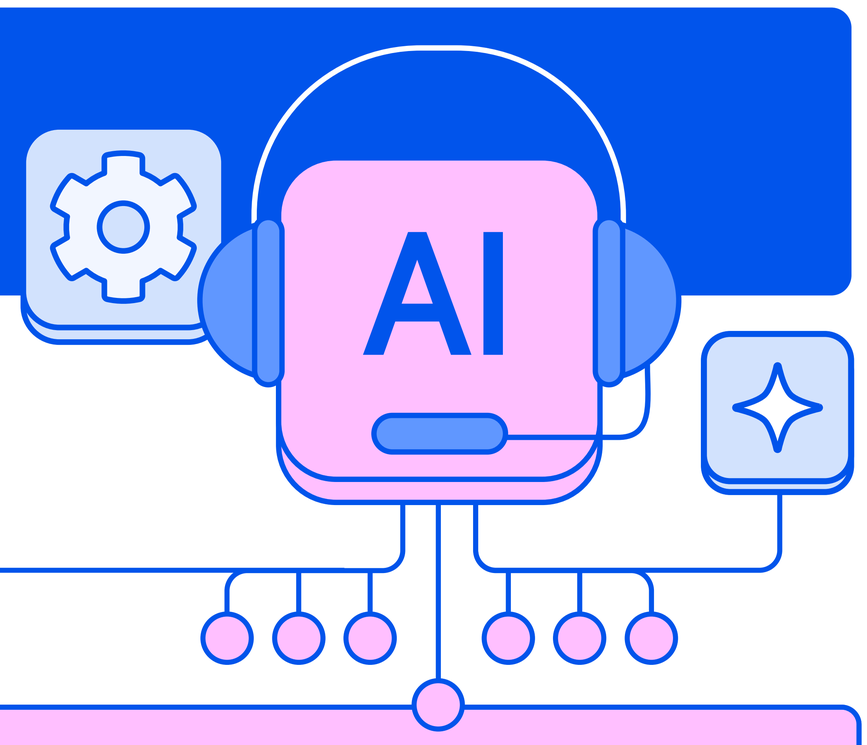
Securing AI Agents 101



Mapping AI in Enterprise

By Use Case

Productivity copilots:
AI assistants embedded in daily work tools.

Internal AI systems:
In-house AI for analytics, knowledge, & employee support.

External AI systems:
AI features in products and services exposed to customers.

By Form

Personal AI (shadow AI)
Unapproved and ungoverned AI tools in use

Enterprise AI:
Official, sanctioned AI assistants deployed across the business.

AI applications
Dedicated AI-powered apps for specific functions or workflows.

Agent architecture
(cross-cutting) Autonomous, multi-tool AI processes that span systems and teams.

Elements of Secure Enterprise AI

Infrastructure
Harden pipelines, configurations, and integrations.

Model
Treat models as intellectual property; protect & monitor them.

Data
Safeguard AI data from leakage or poisoning at every stage.

Application
Threat Model AI apps & agents like modern cloud-native workloads.

Why It Matters

87% of organizations are already using AI services like OpenAI or Amazon Bedrock—yet only 13% leverage AI-specific security posture management (AI-SPM). Read more

Self-hosted models now represent 70% of AI workloads in the cloud. Read more

31% of organizations say lack of security expertise is their biggest AI challenge. Read more

Actionable Steps for Teams

Ask These Four Questions (and act on the answers)

1. **Do I know every AI service, model, and agent running in my environment?** (AI-SPM Inventory)
2. **Do I understand the risks in each AI pipeline?** (misconfig & attack-path analysis)
3. **Can I prioritize which AI risks matter most?** (contextual risk scoring)
4. **Can I detect and respond to AI misuse?** (runtime detection & data scanning)

Spotlight: AI Agents

What is an agent?

An AI agent is an autonomous system that combines tools, data, and logic to execute tasks end-to-end—without step-by-step human direction.

What makes something an agent?

- Autonomous execution: Operates without constant human prompts.
- Multi-step orchestration: Chains tools, APIs, & actions to achieve an outcome.
- Decision-making capability: Chooses next steps based on data or context.

Why agents have become a new surface area for threats

- They act, not just advise: Agents can trigger changes, execute actions, and move data across systems without manual approval.
- They cross boundaries: One agent workflow can span multiple clouds, apps, and data sources, making lateral movement easier for attackers.
- They hold powerful keys: Agents often have broad or persistent permissions that, if compromised, can be abused at scale.

Types of agents in the enterprise

- Productivity agents: Automate tasks like report creation, scheduling, and content workflows.
- Customer-facing service agents: Handle support chats, order processing, and personalization.
- Code & DevOps agents: Write, test, and deploy code and infrastructure changes—often with high-level permissions.

In Wiz's AI research, agent driven architectures have emerged as one of the fastest growing and least governed attack surfaces in enterprise AI.

Top Risks in Enterprise AI

Shadow AI
Untracked AI tools operating outside official oversight, hurting visibility.

Data leakage
Sensitive data exposed by insecure AI workflows and connections.

Excessive permissions
AI models, services, & agents granted overly broad access rights.

WIZ