Silverfort

Silverfort AI Agent Security

Innovate with AI knowing your AI agents and MCP servers are secure. Connect every AI agent to a human identity for accountability and inline access control.

Unlock AI Agent potential, securely

Securely adopt AI by treating AI agents as identities—governed, visible, and protected with the same rigor applied to human users. Simple to deploy so security teams have control, developers can keep building, and innovation can accelerate.

Human accountability for every AI agent

Tie every AI agent action to the initiating human for clear, immutable accountability and auditability.

Inline security controls

Dynamically grant access between AI agents and MCP servers, restricting over-privileged access.

Stop AI agent overreach

Prevent misuse and data leakage. Make lateral movement and privilege escalation impossible.

The AI agent identity security challenge: Neither human nor machine

AI agents exist in the grey area between human and non-human identities, and traditional identity solutions weren't designed for an AI world. AI agents act independently, make decisions, and access sensitive systems, often without oversight. Many organizations adopt the MCP protocol to support these agents, but do so without a clear security framework, exposing critical systems to potential misuse. AI Agents act on behalf of users or services using delegated identities, yet can't always be traced to a human owner, obscuring accountability and increasing the risk of misuse. Agents may have their own access rights—often with high privileges— and their evolving behavior is unpredictable, making it nearly impossible to pinpoint compromise.

This leads to:

- Unattributed or over-privileged tool calls and access
- Prompt mutations through autonomous reasoning with no rollback
- Hidden or indirect prompt injection attacks (XPIA) via user data, web pages, or downstream tools
- Lateral movement across SaaS or cloud services after compromise

Through 2029

>50%

of successful cybersecurity attacks against AI agents will exploit access control issues, using direct or indirect prompt injection as an attack vector.

— Gartner, How to Secure Custom-Built AI Agents, March 2025

Silverfort AI Agent Security: Confidently secure the next wave of AI innovation.

Link AI agents to the human users they represent for accountability to while enforcing access controls inline to prevent data leakage or misuse.

Centralized control

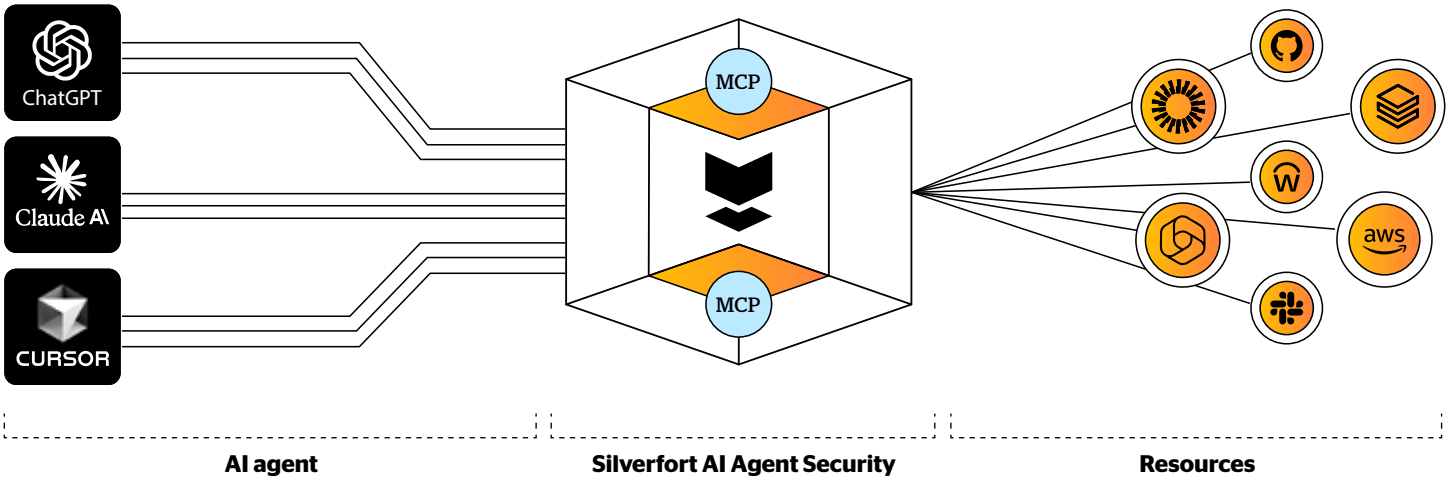
Manage all AI agent configurations, secrets, and policies in a single place.

Dynamically enforce least-privilege

Reduce the risk of unrestricted use with inline access controls.

Quick, non-intrusive deployment

Integrate agents quickly with no impact on end users and developers.



How it works

Silverfort approaches securing AI agents the same way it secures human and service identities: inline and identity-first. Silverfort's model ties every agent's action to a real human owner, enforces least privilege, and logs an immutable audit trail. Silverfort sits inline between the agent and the MCP server, inspecting every call before it reaches the target database, protecting all involved with robust, real-time security controls and preventing misuse and data leakage. The identity-first architecture enables rapid deployment—often within days—to allow immediate protection of AI agent activity.

Combined with precise least-privilege access enforcement, AI agents are strictly limited to actions they are explicitly authorized to perform, so every decision aligns with both security policies and contextual relevance.

About Silverfort

Silverfort secures every dimension of identity—humans or machines across the cloud and on-prem. We deliver end-to-end identity security that is easy to deploy and won't disrupt business operations, resulting in better security outcomes with less work. Discover every identity, analyze exposures, and enforce protection inline to stop lateral movement, ransomware, and other identity threats.