BeyondTrust

AI Agent Security for Endpoints

Enforce least privilege for AI agents running on your endpoints.

Every AI agent is a new kind of non-human identity (NHI) that runs with your users' full privileges and none of their accountability. BeyondTrust AI Agent Security lets you adopt AI everywhere—without creating your next insider threat.

With AI Agent Security, you control the actions of AI agents on your endpoints with policy controls distinct from those you apply to humans. Whether it's a desktop assistant working through local files, an automation built and run by an employee, or a coding assistant like Claude Code, Cursor, or Copilot, you can apply least-privilege principles to limit what those agents can access, execute, and change.

In addition, discover every AI agent and MCP server on every endpoint, including shadow installs. Capture all AI agent activity with per-action attribution and full session transcripts.

"Our Phantom Labs® security research team's controlled research into agentic AI and autonomous AI worms exposed a critical gap: AI agents can act, adapt, and propagate at machine speed. Once an AI agent is given access, permissions, and autonomy, it effectively becomes a new type of identity. Traditional security controls assume a human is behind every decision. That assumption no longer holds. We built AI Agent Security precisely to bring visibility, least privilege, and control to this new class of digital identity."

–Kinnaird McQuade, Chief Security Architect, BeyondTrust

SUPPORTED AGENTS:

GitHub Copilot · Claude Code · Google Cloud · salesforce Agentforce Vibes · CURSOR · Copilot · OpenAI Codex · And more...

- **Visibility: Enable Secure AI Adoption**
  Turn "block it" into a governed "yes," so teams get AI productivity without unmanaged risk.

- **Intelligence: Attribute Every Agent Action**
  Every agent action is logged, and complete transcripts of agent-executed commands provide audit and forensic evidence for compliance requirements such as the EU AI Act, NIST, and ISO 42001.

- **Protection: Mitigate Risk with Controls Designed for AI Agents**
  Restrict AI access to sensitive credentials, limit agents' communication with specific external APIs and MCP servers, and enforce human-in-the-loop approvals before agents act. This helps prevent credential exfiltration, block destructive commands, and ensure EDR has nothing to clean up after the fact.

LEARN MORE beyondtrust.com/products/ai-agent-security
CONTACT US beyondtrust.com/contact