akto

Datasheet

Official Claude Compliance Partner

Akto Atlas for Anthropic

Secure Claude usage across your organization. From Desktop to CLI.

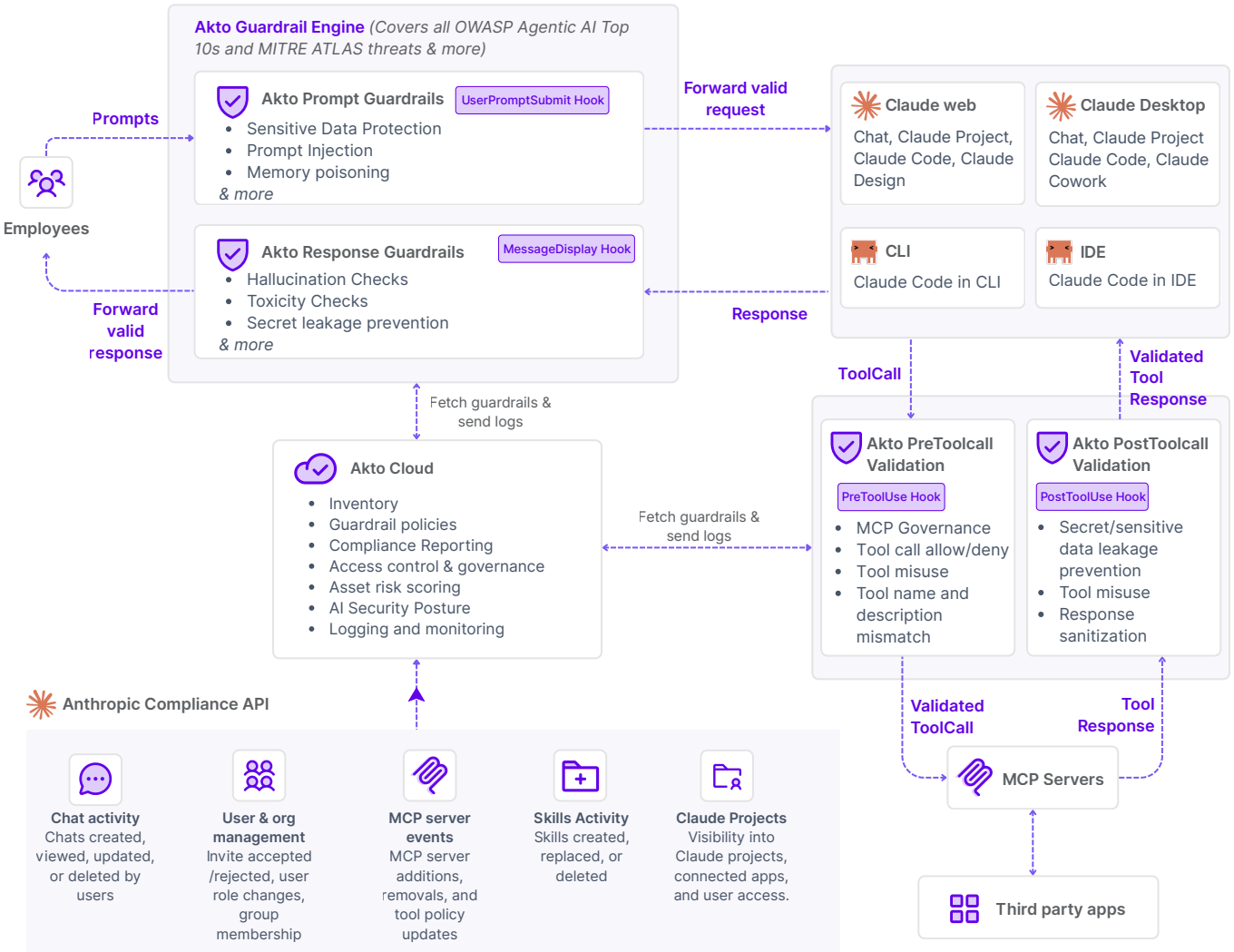
Claude is everywhere in your organization.

- Developers run Claude Code from terminals.
- Employees are building and running skills using Claude.
- Teams collaborate in Claude.ai Projects.
- AI agents connect to internal systems through MCP servers on Claude Desktop.

Every interaction is a potential vector for data leakage, prompt injection, unauthorized access, and shadow AI sprawl. Security teams lack visibility into who is using Claude, what data is flowing through it, and which MCP connections are active across employee devices.

One Security Layer Across Browser, CLI, Desktop, and IDE

Akto Atlas provides end-to-end security for every interaction with Anthropic and Claude.
It discovers Claude usage across browsers, CLIs, and local environments, enforces guardrails on prompts and responses in real time, and inventories MCP connections and agent skills, all without disrupting developer workflows. Atlas connects through browser traffic interception for Claude.ai, and Endpoint Shield plus native hooks for Claude Code, and Claude Desktop.



Use Cases	The Gap	With Akto Atlas
Eliminating Shadow Claude Usage Across the Enterprise	Lack of visibility into who's using Claude, where, or what's being shared. Personal accounts bypass every enterprise control.	Live inventory of every Claude surface, device, session, and user identity. Enforce policies against unmanaged devices, unsanctioned MCPs, and personal account usage. Flag malicious Skills and malicious MCP servers.
Preventing Data Leakage and AI-Specific Threats	PII, credentials, and API keys leak through prompts and responses. Prompt injection goes undetected - traditional DLP is blind to it.	Real-time inspection of every prompt and response. Sensitive data redacted, malicious inputs containing indirect prompt injection are blocked, policy enforced - before the request reaches Claude or the response reaches the user.
Securing Claude Projects, Skills, and MCP Connectors	Claude Projects accumulate sensitive files with consumer-grade visibility. Over a third of audited skills contain security flaws. Malicious skill campaigns have seeded hundreds of poisoned skills. Malicious MCP can exfiltrate data through connectors.	Comprehensive inventory of users, documents, connectors, Skills, and MCP server changes pulled directly from Anthropic's Compliance API into the Akto dashboard. SkillGuard blocks malicious Skill instructions at runtime on the endpoint. Endpoint Shield enforces connector allowlists and inspects connector traffic in real time.
Securing Claude Code and Cowork Autonomous Activity	Claude Code and Cowork operate with high privileges, executing tasks and invoking tools on employee machines. Anthropic's audit logs and Compliance API record what happened. Security teams still need to stop a risky tool call, exposed credential, or unapproved repo access before it runs.	Inline enforcement on every Claude Code and Cowork session. Tool calls checked before they run, credentials intercepted, repo access governed, and every invocation auditable and attributable to a user and device.
Meeting Compliance, Access, and Audit Mandates for AI	No mapping from Claude policy violations to the controls auditors and regulators expect. Native RBAC isn't granular enough.	Audit-ready reports across EU AI Act, NIST AI RMF, MITRE ATLAS, ISO 27001, SOC 2, HIPAA, GDPR, and 15+ standards. Granular RBAC, complete audit trail.

100%	Real-time	Zero	Complete
Claude Surface Coverage	Guardrail Enforcement	Workflow Disruption	Audit trail

Anthropic Suite Connectors

Anthropic Compliance API

Continuously streams Claude chat, user, MCP, and Skills events for org-wide visibility.

Claude Code CLI Hook

Import your Claude Code CLI proxy traffic seamlessly into Akto.

Claude Cowork

Secure your Claude Cowork interactions with Akto guardrails.

Claude Desktop

Secure your Claude Desktop interactions, including Claude Code, with Akto guardrails using Akto hooks.

Claude Inference Hooks

Route every governed Claude prompt through Akto for an allow/deny verdict before inference runs.

Integration Options

Endpoint Shield + Hooks

Deeper enforcement layer for Desktop apps and IDE-based MCP connections, paired with native hook integration. Wraps MCP servers with Akto's security proxy and validates prompts, responses, and tool calls at runtime.

Enterprise EDRs/SASE + Hooks

Deploy at scale via SASE or EDR platforms - Microsoft Defender for Endpoint, SentinelOne, combined with native hooks for prompt, response, MCP and tool call validation at the source. Centralized policy management and unified reporting to the Akto Atlas dashboard.

Anthropic Compliance API Connector

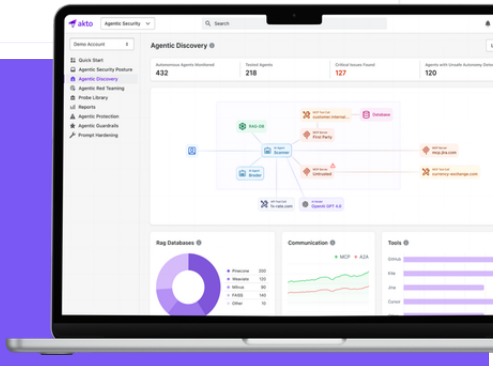
Agentless integration that pulls activity data directly from Anthropic's Compliance API - chat activity, user and org management, MCP server events, and Skills activity. Deploys in minutes with no endpoint installation, and applies asynchronous guardrails to flag policy violations across all discovered Claude.ai usage.

Browser Extension

Lightweight browser extension for rapid, organization-wide coverage of browser-based Claude interactions. Monitors all web activity, file uploads, and API interactions without impacting performance.

Claude Inference Hooks

Inline enforcement inside Claude Enterprise. Anthropic sends every governed prompt from claude.ai, Cowork, and Claude Code to Akto Atlas, which checks it against your Guardrail policies and returns an allow or deny verdict before inference runs. Denied prompts never reach the model.



Protect employee-driven AI usage before it becomes your next attack surface, with Akto.

sales@akto.io | +1 (415) 329-1415