Solution Brief

Netskope Skylight AI Security

Netskope Skylight secures AI across every stage of use. It discovers every app, model, agent, and prompt across cloud, endpoint, and network, then governs and secures that activity from one control plane. With Skylight, security teams can manage risk - and say yes to AI with confidence.

Quick Glance

- Discovers every AI app, model, agent, and prompt across cloud, endpoint, and network
- Extends data loss prevention into AI prompts and agent workflows, redacting sensitive data instead of blocking work
- Authenticates, rate limits, and adversarially tests self-hosted models before and after they reach production
- Governs what AI agents connect to, trust, and do, blocking actions before they execute
- Runs on Netskope NewEdge, in over 80 regions, interconnected with AWS, Anthropic, and OpenAI to inspect AI traffic.

> "94% of organizations report gaps in AI activity visibility."

Netskope 2026 AI Risk and Readiness Report

The challenge

AI now touches data, people, self-hosted models and autonomous agents, often before security teams know it exists. Only 6% of organizations have complete visibility into their AI pipeline.

Gaps in AI visibility are reported by 94% of organizations, and 91% say they cannot stop a risky agent action before it executes. Nearly half (47%) of users access AI through personal accounts that bypass corporate controls, and 61% report unmanaged AI tool use.

Employees paste sensitive data into public AI apps, developers connect coding agents to unreviewed servers, and self-hosted models go live without adversarial testing. Traditional tools inspect one layer at a time, leaving no single view of what data, people, models, and agents touch.

The solution

Netskope Skylight delivers unified AI security through the Netskope One platform. Discover every user, app, agent, and interaction across SaaS, cloud, endpoint, and network, with risk intelligence fed to Netskope's single AI control plane: the Netskope Skylight AI Command Center. That intelligence drives the Zero Trust Engine, enforcing realtime controls over people, applications, agents, and data. Guardrails, threat, and data protection secure interactions inline. With NewEdge and AI Fast Path, security teams discover, govern, and secure AI at speed.

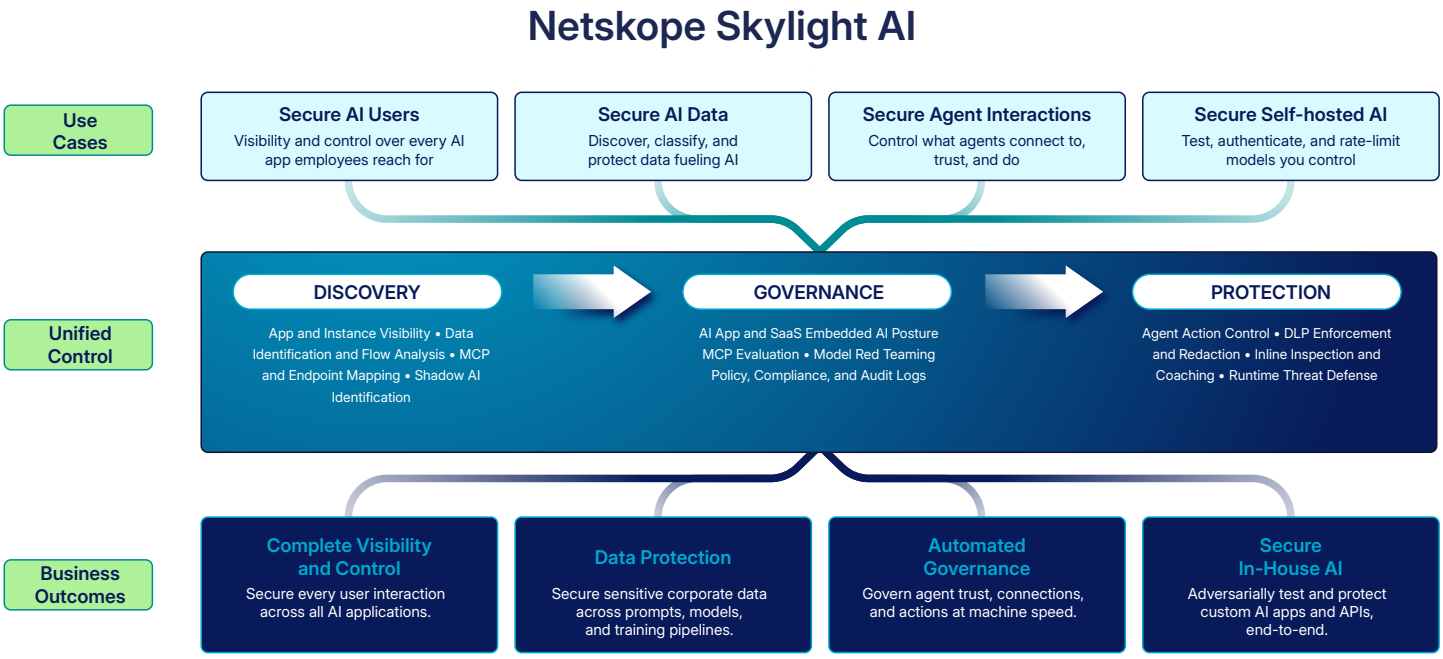
Empower safe AI adoption by users

Employees rarely set out to cause breaches, but they can cause them in the process of trying to get their job done efficiently. Shadow use follows people onto their work-issued devices, where a local model or browser extension runs outside the network's view. A security block with no explanation just moves the same behavior elsewhere.

Netskope One Next Gen Secure Web Gateway (NG-SWG) and Netskope One CASB (cloud access security broker) inspect every prompt and response inline across more than 1,800 AI apps, distinguishing between corporate and personal accounts. Netskope One Client, which is already on managed endpoints in the same environment, extends that visibility to local models, browser and integrated development environment (IDE) extensions, as well as loopback connections that never touch the network.

Security controls operate at account - not app - levels, reducing workload for security teams: approving a tool for one account clears it for use across entire user groups if appropriate. And where the moment calls for education, instead of blocking use of a tool, Netskope coaches the employee on the risk, highlights corporate policy, and even steers the user toward the company approved option in real time, explaining the risk instead of leaving them to guess.

> Control operates at the account, not app level, so approving a tool does not mean approving shadow use.



Secure self-hosted private AI before and after launch

When an organization deploys its own model, or an application that calls it directly, it owns the entire threat surface. No public AI provider hardens that endpoint. Functional testing proves a model works, not that it survives an attack, and jailbreak attempts succeed about 20% of the time against untested models, in as few as five interactions. Once live, traffic comes from applications and agents rather than a human, and an unauthenticated call can spike compute costs overnight before anyone notices.

Part of Netskope Skylight, AI Red Teaming automates adversarial testing against models an organization fully controls, drawing on more than 18,000 scenarios, including multi-turn jailbreaks such as skeleton key and crescendo attacks. It runs in continuous integration and continuous delivery (CI/CD) pipelines before every release and after. Ensuring models are tested against advanced threats before attackers strike.

Once a model clears red team tests, AI Gateway authenticates every call with unique tokens, enforces model-level policy, rate limits usage against runaway cost, and logs every request for audit, whether the model runs on-premises or in a hyperscaler environment. AI Guardrails then blocks the same prompt injection and jailbreak techniques in production at runtime, so the model stays hardened under real load, not just at launch. That combination keeps defenses current after launch.

Govern what AI agents connect to and do

Once an agent reaches past infrastructure an organization controls, a public MCP (Model Context Protocol) server or another company's tool, that organization is trusting something it did not build and cannot fully inspect. 91% of organizations cannot stop a risky agent action before it executes, and 88% reported a confirmed or suspected agent security incident in the last year. An agent that deletes a repository, provisions infrastructure, or forwards a file does it through a normal application programming interface (API) or MCP call, without waiting for sign-off.

Agentic Broker risk-assesses MCP servers before an agent connects, public or private, using the Netskope dynamic and continuously updated risk scores to surface risk before the connection completes. It decodes every MCP session in real time and integrates with the same Netskope data loss prevention and access policy that already governs people.

Agent Action Control classifies every agent action into one of nine intent-based categories, including data destruction, credential manipulation, and remote code execution, and assigns a risk level based on what the action touches. Security teams build a profile that decides to block, allow, or alert, and attach it to specific agents, so a coding assistant and a chat app follow different rules. When a coding agent attempts to delete a production repository instead of a test branch, the action is blocked before it can execute.

Secure the data that fuels every AI interaction

AI cannot run without data. That data moves through prompts, responses, agent retrieval, and the pipelines behind model training and retrieval-augmented generation (RAG). Only 6% of organizations have complete visibility into their AI pipeline, and the average organization sees 223 AI data policy violations a month. Training data creates exposure long before a prompt does. Most exposure starts earlier, in data stores nobody discovered or classified before they fed a training set or RAG index.

Netskope treats every product that touches AI data, including Next Gen Secure Web Gateway, cloud access security broker (CASB), Agentic Broker, AI Gateway, data loss prevention (DLP), and endpoint discovery, as a sensor feeding one control plane, AI Command Center. Before data reaches a prompt, training run, or RAG index, data security posture management (DSPM) discovers and classifies it across infrastructure as a service, platform as a service, on-premises, and Software-as-a-Service (SaaS) stores, including shadow data stores. Pre-trained classifiers provide speed and Train Your Own Classifiers (TYOC) identifies sensitive data unique to the organization. Analytics trace that data from user to AI app and instance.

The same data security policy protecting email, endpoint, and SaaS extends into prompts, responses, and training pipelines, blocking shadow AI entirely, or redacting only sensitive data so work can continue. This means a user summarising a draft report with an AI app still gets their summary, while sensitive information stays under the firm's control, and never reaches the AI model.

> Unified data security covers email, files, AI prompts, and training pipelines, so sensitive data stays protected wherever it moves across the business.

BENEFITS	DESCRIPTION
Discover AI, everywhere	Builds one continuously updated inventory of every AI app, model, agent, and workload across cloud, endpoint, and network, replacing multiple dashboards with a single view.
Protects data before it ever trains a model	Discovers and classifies data with DSPM before it reaches a prompt, training run, or RAG index, then extends data loss prevention policy from email and files into AI prompts and responses, redacting sensitive data instead of blocking the task, or blocking the upload and advising the user.
Tells corporate accounts from shadow AI	Distinguishes managed accounts from personal ones across more than 1,800 AI apps, so control operates at the account level, not just the app level.
Adversarially tests models before launch	Runs more than 18,000 private attack scenarios against self-hosted models in continuous integration and continuous delivery pipelines, retesting every finding before release.
Authenticates every model and agent call	Enforces unique tokens, model-level policy, and rate limits on every call to a self-hosted model, on-premises or in a hyperscaler environment.
Risk-assesses MCP servers before connection	Evaluates public and private Model Context Protocol servers with dynamic and continuously updated risk scores, surfacing weak authentication before an agent ever connects.
Classifies every agent action by intent	Assigns each agent action a risk level and applies a block, allow, or alert policy per agent, so one agent's rules do not carry over to another.
Correlates every signal in one place	Feeds every sensor, from endpoint to cloud to network, into AI Command Center, ranking real exposure instead of surfacing the loudest alert.
Runs at AI speed without added latency	Delivers inspection and enforcement over Netskope NewEdge, spanning over 120 data centers across more than 80 regions, peered directly with providers like OpenAI and Anthropic. AI Fast Path reroutes traffic around congestion, keeping AI calls under five milliseconds round trip in nine out of ten cases.
Netskope AI Risk AISecOps Agent speeds up triage	A companion capability, not part of Skylight, that pulls full context from every Skylight sensor, the asset, the data touched, and the identity involved, into one case, scores severity, with clear recommended actions for analysts.