Solution Brief

# Netskope One AI Security for Federal Agencies

Federal agencies must rapidly adopt AI while meeting strict mandates for security, transparency, and risk management. However, AI introduces new risks across data, applications, and autonomous systems. The Netskope One platform enables secure AI adoption by governing the full AI lifecycle, while protecting sensitive mission data in real time.

### Quick glance

- Gain unified visibility and control across all AI activity: users, SaaS applications, models, and autonomous agentic workflows
- Protect sensitive data across the full AI lifecycle, including prompts, outputs, and model interactions using semantic DLP
- Govern AI behavior with the Netskope Zero Trust Engine, ensuring continuous verification and least privilege
- Detect and block AI-specific threats, including prompt injection and jailbreaks
- Align AI adoption with OMB, NIST AI Risk Management Framework (AI RMF), and CISA guidance.

> "While AI is improving operations and service delivery across the Federal Government, agencies must effectively manage its use."
>
> Office of Management and Budget, M-24-10

## The challenge

Federal agencies are accelerating AI adoption under mandates such as OMB M-24-10 and the NIST AI Risk Management Framework. However, AI introduces new risks including sensitive data exposure in genAI tools, embedded AI within SaaS applications, and autonomous agents capable of taking actions across systems.

Traditional security tools lack visibility into AI usage, cannot inspect prompts or model interactions, and are not designed to govern AI behaviors or enforce policy across AI-driven workflows. This creates blind spots across the AI lifecycle, from data input to model output to autonomous action, making it difficult for agencies to manage risk, ensure compliance, and confidently scale AI across mission environments.

**Result:** AI innovation is outpacing the ability to secure and govern it.

## Netskope One AI Security for federal agencies

Netskope enables federal agencies to securely adopt AI by delivering unified visibility, control, and governance across the full AI lifecycle. Using the FedRAMP High authorized Netskope NewEdge Government platform, agencies gain real-time visibility, inline enforcement, and zero trust-based controls across users, SaaS applications, models, and autonomous agents.

Netskope secures AI interactions, enforces guardrails, and validates systems continuously. This enables safe, compliant AI adoption without increasing risk, adding latency, or slowing the mission.

## Secure the full AI lifecycle with unified control

Federal agencies must secure AI across every stage, from data input, to model interaction, and from output to autonomous action.

Netskope One AI Security provides unified lifecycle governance across all AI activity, securing both human and non-human interactions.

- Netskope One AI Gateway governs how internal applications and APIs interact with privately hosted or public AI services, providing real-time visibility, inline enforcement, and data protection across all app-to-LMM traffic.
- Netskope One Agentic Broker governs how AI agents operate, decoding Model Context Protocol (MCP) traffic to control what they can access, what actions they can take, and how they interact with tools and data sources.
- Netskope One AI Guardrails provides AI-specific runtime protection that detects and blocks prompt injection, data manipulation, and adversarial AI behavior in real time.

**Result:** Agencies gain full visibility and control of AI, from input to action, reducing risk and eliminating blind spots across the entire AI ecosystem.

> "Data security is a critical enabler that spans all phases of the AI system lifecycle. ML models learn their decision logic from data, so an attacker who can manipulate the data can also manipulate the logic of an AI-based system."
>
> CISA, NSA and the FBI, AI Data Security Best Practices

## Discover and govern AI usage across the enterprise

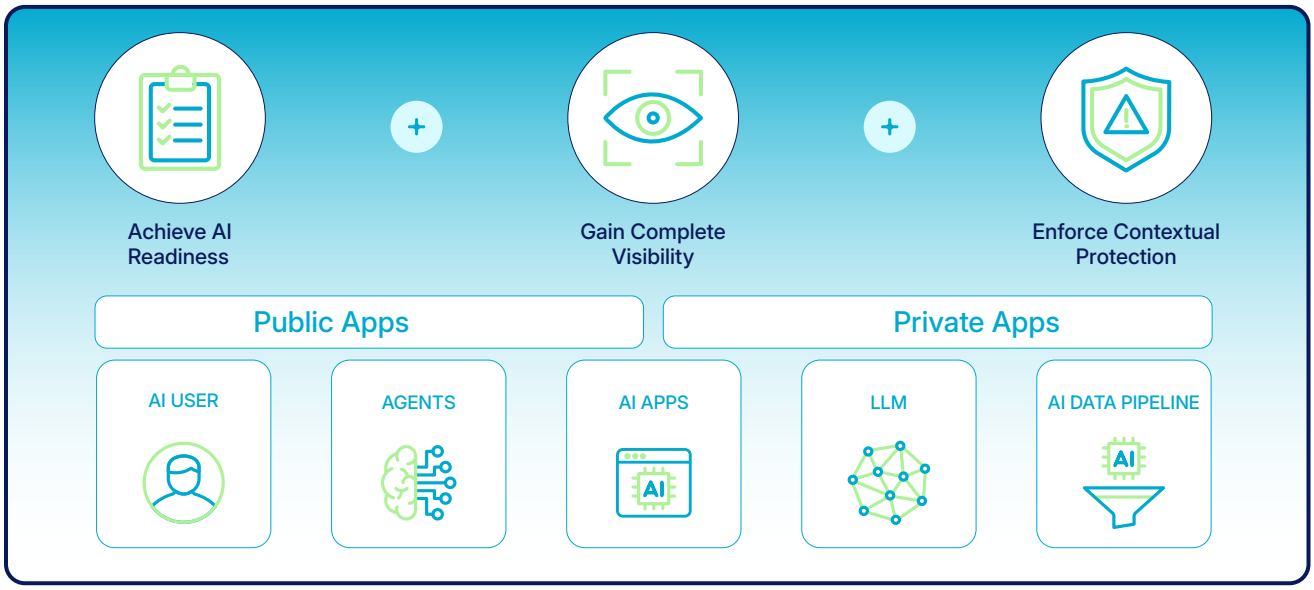
Agencies must maintain continuous visibility into how AI is being used across the mission, including AI embedded within enterprise applications.

Netskope One AI Security provides:

- Discovery of generative AI tools, embedded AI in SaaS platforms, and unmanaged AI usage
- Classification of AI applications by risk, data sensitivity, and mission impact
- Visibility into user prompts, interactions, and AI-generated outputs, utilizing real-time "coach and pivot" alerts to safely guide users toward approved tools
- Continuous monitoring aligned to federal AI inventory requirements
- Identification of unmanaged or managed AI tools

**Result:** Agencies gain full visibility and control of AI, from input to action, reducing risk and eliminating blind spots across the entire AI ecosystem.

> NIST's 2025 red-team exercises found that novel attacks against AI agents succeeded 81% of the time.




## Prevent sensitive data exposure with Netskope One AI Guardrails

Federal guidance requires agencies to protect sensitive data, privacy, and mission-critical data across all AI usage.

Netskope One AI Guardrails enforces real-time protections across the lifecycle by:

- Blocking or warning on sensitive data in prompts before submission
- Redacting or tokenizing data before it reaches AI models using semantic data loss prevention (DLP) that evaluates the actual meaning and intent of data, preventing exposure even if an AI rewrites or transforms the content
- Inspecting AI outputs for sensitive or policy-violating content, providing real-time content moderation to block AI-specific threats like prompt injections, jailbreaks, and adversarial misuse
- Enforcing consistent controls across SaaS, web, APIs, and private AI tools and autonomous agentic workflows

Combined with Netskope One DLP, agencies extend zero trust data protection directly into AI workflows, ensuring least-privilege data usage at every stage without slowing down the mission.

**Result**: Sensitive data is protected continuously across prompts, models, and outputs, delivering total data control for the enterprise.

## Validate AI systems with continuous Netskope One AI Red Teaming

Federal agencies must ensure privately hosted AI systems are secure, reliable, and resilient before and after deployment.

Netskope One AI Red Teaming enables proactive and continuous validation of AI systems by:

- Simulating adversarial prompts and real-world misuse scenarios using an automated library of over 18,000 test cases and seed prompts
- Identifying prompt injection, model manipulation, and data leakage risks, mapping all findings against industry standards like OWASP Top 10 for LLMs
- Detecting unsafe, biased, or noncompliant outputs
- Continuously testing AI systems as models evolve over time by integrating stress tests directly into CI/CD pipelines via APIs
- Validating alignment with agency policies and risk thresholds

Combined with Netskope One DLP, agencies extend zero trust data protection directly into AI workflows, ensuring least-privilege data usage at every stage without slowing down the mission.

**Result**: AI systems are continuously validated against real-world attack techniques, reducing operational and mission risk before vulnerabilities ever reach production.

## Secure and govern AI agents and autonomous systems

As agencies adopt AI agents that can take action, risk extends beyond human outputs to system-level, autonomous behavior.

Netskope One Agentic Broker provides governance over AI-driven actions by:

- Decoding and securing MCP traffic enforcing policy on what AI agents can access and execute
- Controlling AI-to-tool, AI-to-API, and AI-to-data interactions
- Applying least-privilege access and context-based zero trust policies to non-human identities
- Monitoring autonomous decision-making and behavior to detect anomalies or potential tool poisoning
- Integrating with Netskope One DLP to prevent sensitive mission data or credentials from being exfiltrated during agentic interactions.

**Result:** Agencies control not only what AI produces, but what AI is allowed to do across the enterprise.

## Enforce real-time AI governance with the Netskope Zero Trust Engine

AI governance must be continuous, adaptive, and enforceable across all environments.

Netskope One AI Security enables:

- Inline enforcement of AI policies across all human and non-human interactions, utilizing the high-performance Netskope NewEdge Network to eliminate the AI security latency tax
- Context-aware controls based on identity, device, application risk, and behavioral anomalies
- Real-time blocking, coaching, or step-up actions
- Continuous monitoring and adaptive risk scoring for both managed and unmanaged AI tools
- Policy enforcement aligned to zero trust principles of continuous verification and federal mandates like the NIST AI Risk Management Framework (AI RMF)

**Result:** AI governance becomes dynamic, real-time, and fully enforceable, enabling agencies to accelerate AI adoption without increasing mission or data risk.

| BENEFITS | DESCRIPTION |
|---|---|
| Secure every AI interaction | Gain total visibility and control across all AI traffic, including humans, internal applications, and agents, through a unified, high-performance layer. |
| Reduce AI data exposure risk | Prevent sensitive data leakage across prompts, outputs, and model interactions. |
| Control AI-driven actions | Govern what AI agents can access, execute, and automate across systems. |
| Detect AI-specific threats | Identify and block prompt injection, manipulation, and adversarial AI behavior in real time. |
| Meet federal AI mandates | Align EO, OMB, NIST AI RMF, and CISA guidance for governance, risk, and oversight. |
| Eliminate shadow AI | Discover and control unmanaged AI usage across the agency |
| Enforce zero trust for AI | Extend identity, context, and data-aware controls to every interaction and action utilizing the Netskope Zero Trust Engine, ensuring continuous verification across the entire AI ecosystem. |
| Harden AI systems before deployment | Continuously test and validate AI models against adversarial scenarios pre and post deployment. |

netskope

**Interested in learning more?** Request a demo

Netskope (NASDAQ: NTSK), a leader in modern security and networking for the cloud and AI era, addresses the needs of both security and networking teams by providing optimized access and real-time, context-based security for the AI ecosystem inclusive of agents, applications, tools, LLMs, people, devices, and data. Thousands of customers, including more than 30 of the Fortune 100, trust the Netskope One platform, its Zero Trust Engine, and its powerful NewEdge network to reduce risk and gain full visibility and control over cloud, AI, SaaS, web, and private applications – providing security and accelerating performance without trade-offs. Learn more at netskope.com, Netskope.ai, on LinkedIn, and Instagram.