netskope

Securing AI: 5 Crucial Conversations for CISOs

Contents

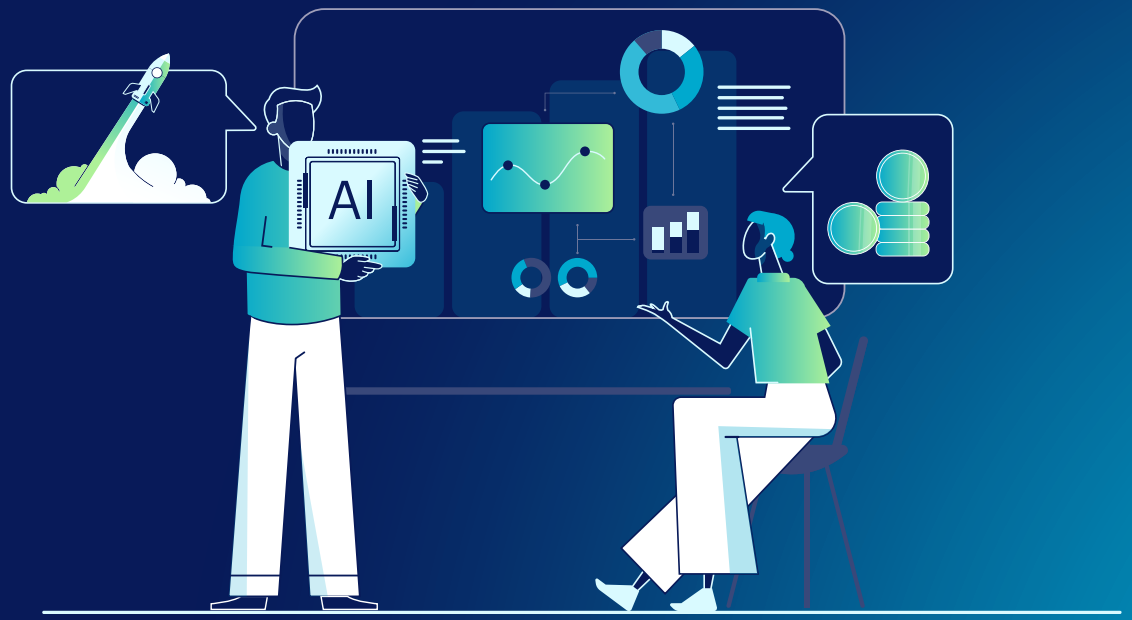
Introduction: A dual mandate for AI

The high profile of technology in today's modern organization has put the IT department in the spotlight like never before. As a result, IT leaders find themselves engaging in a series of crucial conversations with their CEO, board, and senior leadership peers as they seek to establish the ways in which technology can best contribute to business success—and no topic is more prominent in those discussions than AI.

AI poses a particular conundrum to CIOs, CISOs, and their teams. In Netskope's Crucial Conversations report[1], we revealed that CEOs are giving their IT chiefs a double-edged mandate: integrate AI to encourage experimentation and drive measurable business value, but also cut costs, act as gatekeepers against overspending, avoid hype, and protect against potential data leaks or security breaches.

In short, IT leaders must use AI to enable disruptive innovations, while simultaneously defending against the risks that it brings. It's a duality that heaps significant pressure onto their shoulders.

Every organization is at a different stage in AI maturity. Some are still identifying the use cases where AI can make an impact, while others are motoring ahead, building their own AI apps and encouraging staff to widely adopt these tools. Everyone is racing to leverage AI to accelerate their growth, but they're moving at different speeds from different starting points.



[1] https://www.netskope.com/crucial-conversations

Five steps to AI adoption

Wherever your enterprise is in the AI maturity curve, there are vital security considerations that need to be taken on board. The key is to plan your security strategy with a thorough understanding of the risks at each stage.

1. **Can we experiment with AI tools**, while managing the risks of shadow AI?
2. **Are we able to leverage embedded AI in SaaS platforms** without enabling unapproved data sharing?
3. **How do we manage stand-alone AI apps** and stop data leakage?
4. How can we avoid harmful or biased model responses and common application vulnerabilities when **building private AI apps**?
5. How do we avoid granting over-permissive access when **deploying autonomous agents**?

Experimentation

Integrated AI in SaaS platforms

Sanctioned stand-alone AI apps

Private AI apps

Autonomous agents

Toward more productive C-suite conversations

AI is not just the main topic of conversation in tech circles today, it's also one of the key priorities among C-suites and boardrooms. From our research with CEOs [1], we know that they are enthusiastic about AI's potential and want their IT leaders to create a pathway for adoption and integration where appropriate, without falling for industry hype.

The challenge for IT practitioners, especially in the security function, is to deploy AI in a way that avoids trade-offs between performance and security, and that stays compliant while reducing cost and complexity. Likewise, as implementation plans become more granular, it is critical that the business benefits and risks remain in view.

With this eBook, we hope to contribute toward these outcomes. ***Securing AI: 5 Crucial Conversations for CISOs*** is intended to help security teams move forward with confidence in their AI journeys, holding more productive conversations about AI's challenges and opportunities with colleagues. The ultimate aim is to help organizations embed AI-ready principles into their security strategy, and security principles into their AI strategy—turning security into a driver of growth, rather than a blocker.

Disrupting with AI, defending against risk—three principles to prioritize.

Wherever you are in your AI adoption journey, three principles apply to securely harness the technology's potential.

1. **Visibility.** Security teams need total visibility over their AI landscape, so they know which AI tools are being used, and how.
2. **Protection.** Practitioners need to enforce contextual protection so that dynamic, adaptable security safeguards the business without holding back innovation.
3. **Readiness.** By proactively understanding their data and applications, security professionals can achieve AI readiness that sets their organization up for success.

[1] *https://www.netskope.com/crucial-conversations*

Step 1: Experimentation

Can we experiment with AI tools, while managing the risks of shadow AI?

When ChatGPT launched in November 2022, it took the world by storm—and enterprises by surprise. Almost immediately, employees began using personal instances of AI chatbots to help speed up or solve their work tasks. Today, shadow AI remains an ongoing issue for many organizations: According to Netskope Threat Labs research[1], a remarkable 72% of enterprise users are still using personal accounts to access ChatGPT, Google Gemini, and other popular genAI apps at work in 2025.

This issue is only getting more complicated. Established SaaS apps almost all include embedded AI functionality, AI models are now communicating with each other directly, agents can be created through natural language so they are no longer the preserve of the technically competent, and all of these instances of AI are interacting with more data and applications than a human ever could. As a result, shadow AI is expanding at an unprecedented rate.

72% of enterprise users are still using personal accounts to access genAI apps at work.

Netskope, Generative AI Cloud and Threat Report 2025

[1] *https://www.netskope.com/resources/reports-guides/cloud-and-threat-report-generative-ai-2025*

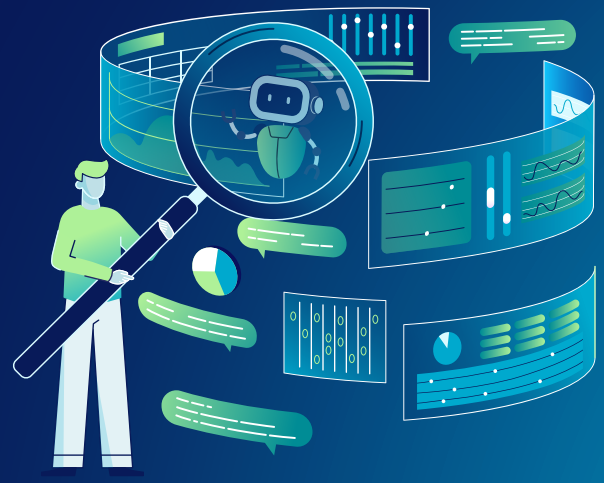
Security teams urgently need both breadth and depth of insight into their AI landscape. They must ensure they have visibility across the whole range of AI tools used within their organization, including unmanaged apps and personal instances. They also need to go deep and understand what users and agents are doing within those interactions.

Only that level of in-depth insight will enable security teams to move beyond blind trust and gain strategic control over their organization's AI activities.

The brutal reality is that you can't secure what you can't see.

How we do it

Through Netskope One Cloud Access Security Broker (CASB) and Next Generation Secure Web Gateway (NG-SWG), enterprises can achieve granular visibility into AI activities within their environment. Our AI dashboard provides granular detail on which users are accessing which applications, as well as which action they're performing. It also helps enterprises perform inline discovery and inspection of all public LLM interactions (data in motion), including user-to-app interactions.



Step 2: Embedded AI in SaaS platforms

Are we able to leverage embedded AI in SaaS platforms without enabling unapproved data sharing?

LLMs and dedicated AI apps are no longer the only risk vectors from an AI perspective. As the technology evolves, AI capabilities are being built into more and more SaaS applications—from video calling platforms to productivity tools and sales management systems.

These SaaS tools are often deeply embedded within modern enterprises, which rely on them for key daily functions of the business, making them almost impossible to block or remove. And AI features typically speed up productivity in a way no organization would want to quash.

New AI functionality is often added with minimal friction—for instance, included in a general update with very little information provided about the data usage terms and conditions. A video calling app, for example, could turn on an AI notetaker by default, recording and storing company-sensitive information. This can easily catch security teams unaware.

An existing SaaS app could introduce new AI capabilities—and even automatically enable them—potentially catching security teams unaware.

Security professionals need to stay on top of their SaaS application landscape, with visibility of AI features, what they do, and the associated contractual terms and conditions relating to data governance. This should include understanding factors such as how each application uses AI, whether it uses your data to train its models, whether it complies with key regulations, and whether its AI features can be disabled.

Organizations should also strongly consider categorizing and classifying sensitive data in order to be able to enforce specific policies that protect company IP or regulated data (for example) while allowing looser rules around non-sensitive information.

How we do it

Netskope's Cloud Confidence Index (CCI) is a repository of over 85,000 SaaS applications, providing rich risk context and empowering security teams to make informed decisions about which AI-enabled apps to allow, restrict, or block.

Step 3: Managed stand-alone AI apps

How do we manage stand-alone AI apps and stop data leakage?

By now, many organizations have selected their preferred AI tool, such as OpenAI's ChatGPT, Microsoft's Copilot, Google's Gemini, or Anthropic's Claude. Standardizing around one system company-wide brings obvious advantages in terms of enterprise-ready capabilities, reinforced learning, and security protections. And if the organization also then blocks other AI systems, it narrows the potential attack surface too.

However, that approach doesn't remove risk entirely. A corporate AI tool can only become truly valuable if it links to other documents and information sources inside the enterprise. That could still enable individual users to pull data from internal documents they shouldn't have access to—in other words, causing data leakage inside your organization.

Someone working in the marketing department could ask an overly-permissioned enterprise AI about upcoming features on the product roadmap, and be served information pulled from confidential documents they are not supposed to access.

How we do it

Netskope actively guards against AI-specific threats during runtime. If a user attempts to input sensitive information, Netskope One Data Loss Prevention (DLP) instantly intervenes, blocking PII, source code, or proprietary secrets from entering the AI model. This can also trigger a coaching pop-up to educate the user.

Simultaneously, Netskope One AI Guardrails provides real-time content moderation for every interaction. It analyzes the intent behind prompts and responses to automatically block sophisticated malicious attacks, such as prompt injections and jailbreak attempts. Furthermore, Guardrails enforces responsible AI usage by filtering harmful or discriminatory content and blocking the delivery of copyrighted materials. By combining these DLP and Guardrail capabilities, organizations can proactively coach users while securing the entire AI ecosystem from data leaks and emerging threats.

Step 4: Private AI apps

How can we avoid harmful or biased model responses and common application vulnerabilities when building private AI apps?

Organizations in highly regulated industries (i.e. healthcare, financial services, government) are leading the way in developing private AI applications. As their confidence with AI grows, many are turning to locally run models trained on an organization's own data to reduce risks around data residency, privacy, compliance, and third-party exposure, while improving relevance and reliability.

A third of organizations are already using OpenAI services via Azure, 27% use Amazon Bedrock, and 10% are building on Google Vertex AI[1]. These enterprise-grade platforms all offer secure, cloud-based AI services that provide stronger privacy controls and deeper integration options than their public versions.

However, building private AI also shifts security responsibility to the organization. Beyond runtime protection against AI-specific threats and misuse by employees, an additional attack surface lies in the tools used to design and deploy these systems, which may lack built-in safeguards.

An important consideration when operating an AI model on-premises is whether it's susceptible to vulnerabilities. If an organization customizes an open source model, for example, the security team should still rigorously test the code and ensure no malicious components have been introduced—for example, code that could capture or transmit prompts to an external source.

Another consideration is to confirm that the organization's training data does not contain any unwanted information. Teams should check for anything biased, sensitive, or harmful in these often very large datasets.

How we do it

By centralizing authentication, traffic management, and content inspection between private apps and LLMs, Netskope One AI Gateway ensures autonomous agentic data flows remain governed and secure. Additionally, Netskope One AI Red Teaming proactively stress tests custom models by automating adversarial simulations within CI/CD pipelines to uncover vulnerabilities like prompt injections.

Netskope One AI Guardrails mitigates sophisticated attacks—including prompt injection and jailbreak attempts—through real-time analysis of all traffic, while also serving as a content moderator to identify and control harmful or discriminatory content for both human and agentic interactions.

In addition, with Netskope One DSPM, security leaders can gain visibility into—and control over—their data, wherever it is. This helps them discover and classify sensitive information, for example, that might be used to train an AI model.

[1] *Netskope Threat Labs, Netskope Cloud and Threat Report 2026*

Step 5: Autonomous agents

How do we avoid granting over-permissive access when deploying autonomous agents?

Agentic AI is the latest favorite child in the hype of AI. Indeed, many commentators have proclaimed it as a key part of the future of enterprise technology, and analyst firm Gartner® predicts that, by 2028, at least 15% of daily business decisions will be made autonomously through agentic AI, up from 0% in 2024[1].

While deployment of this technology is still in its early stages, Netskope Threat Labs research from August 2025 discovered a critical mass of users across organizations already either building AI agents or leveraging the agentic AI features in SaaS solutions.

For example, GitHub Copilot is now used in 39% of organizations, and 5.5% have users running agents generated from popular AI agent frameworks on-premises. According to the researchers, 66% of organizations have users making API calls to api.openai.com and 13% to api.anthropic.com[2].

39% of organizations use GitHub Copilot and 5.5% run AI agents generated from popular frameworks on-premises.

Netskope, Cloud and Threat Report: Shadow AI and Agentic AI 2025

[1] *Gartner press release, Gartner Identifies the Top 10 Strategic Technology Trends for 2025, October 21, 2024*

[2] *https://www.netskope.com/resources/cloud-and-threat-reports/cloud-and-threat-report-shadow-ai-and-agentic-ai-2025*

At the moment, many businesses don't have a firm grasp on the extent of their agentic AI estate. With this field moving so fast, and with new functionality added all the time, shadow agentic AI is an increasingly significant facet of the overall shadow AI problem.

As adoption of AI agents grows—along with the breadth of their capabilities across the organization—the security risks that come with them will multiply. It will be imperative that teams understand the action taken by each agent and put appropriate controls and policies in place to manage permissions and activities performed.

AI-powered apps depend upon authenticated communication between internal applications, autonomous agents, and privately hosted LLMs. For this they use Model Context Protocol (MCP) and APIs, but while the protocols themselves are secure communication routes, these non-human interactions open up a critical security blind spot. APIs and MCP allow AI agents to interact directly with sensitive data and tools, bypassing traditional human-centric security. This gap creates the risk of autonomous interactions without oversight, leading to credential leaks, malicious tool poisoning, and unauthorized data exfiltration.

How we do it

Netskope One AI Gateway operates as a software-defined gateway to intercept and govern API traffic between internal applications, autonomous agents, and privately hosted LLMs, ensuring only authenticated agents can communicate with LLMs by requiring a valid, AI Gateway-generated token for every request.

Netskope One Agentic Broker provides unified visibility and real-time protection for MCP enabled applications—including AI code editors, chat interfaces, and developer tools—by decoding and securing MCP traffic between AI agents and data sources: bridging the gap between human-to-LLM interactions and machine-to-machine AI workflows. This ensures a consistent security posture that protects sensitive corporate data while enabling the speed and scale of agentic automation.

Conclusion: Managing AI risks without trade-offs

AI provides an era-defining challenge and opportunity for CIOs, CISOs, and their teams. It binds them more closely to business strategy and growth than ever before, amplifying their influence and impact. But it also poses significant and rapidly evolving risks to organizational data, revenue, and reputation—upping the stakes of every breach and hack.

To navigate this landscape, IT leaders need a clear framework for understanding AI's disruptive possibilities and defensive requirements—giving them the confidence to talk about AI with their non-technical colleagues, in order to manage AI's potential risks and adhere to rigorous compliance standards.

For today's IT and security practitioner, the key is to secure AI adoption end-to-end to empower safe innovation, while maintaining resilient business operations.

Netskope One is a single, consolidated platform that provides a way to manage AI's risks without performance or user experience trade-offs, simultaneously reducing complexity and ensuring compliance.

As more enterprises move through their AI adoption journey—from initial experimentation to agentic deployment—IT leaders can play a key role in promoting and enabling business innovation. By safely unlocking the benefits of AI, today's CIO and CISO can drive business impact that moves their organization to the next level.

Netskope One AI Security provides a single solution to govern your AI ecosystem and protect your data. It secures users and automated agents across public SaaS, private AI tools, and agentic workflows. Combining high-performance with context-aware zero trust controls, Netskope enables organizations to move on from AI experimentation to unlock AI advantage.

Read more about what CEOs want from their IT leaders in Netskope's Crucial Conversations report here.

About Netskope

Netskope is a leader in modern security, networking, and analytics for the cloud and AI era. The unique architecture of its Netskope One platform enables real-time, context-based security for people, devices, and data wherever they go, and optimizes network performance—without trade-offs or sacrifices. Thousands of customers and partners trust the Netskope One platform, its patented Zero Trust Engine, and its powerful NewEdge Network to reduce risk, simplify converged infrastructure, and provide full visibility and control over cloud, AI, SaaS, web, and private application activity.

Interested in learning more?

Request a demo