netskope

Cybersecurity INSIDERS

2026

AI Risk and Readiness Report

Overview

For most organizations, this is the year AI becomes infrastructure. Agents now execute actions autonomously: modifying records, creating accounts, and pushing code through API calls that complete before any human reviews them. That makes every AI deployment a security risk, whether organizations treat it as one or not.

The security stacks found in most organizations today were built for a different world: one where humans were the only actors, processes were deterministic, data stayed in recognizable forms, and trust was verified at the browser. That world no longer exists.

This report is based on a comprehensive survey of 1,253 cybersecurity professionals, exploring the ways organizations are securing AI, with consideration for governance, visibility, data protection, and agent control.

Key Findings:

- **Adoption of AI has outpaced security governance**
  AI tools are now deployed at 73% of organizations surveyed, but governance that enforces security and policy in real time has reached only 7%. That leaves a 66-point structural deficit, which is widening as AI adoption continues to accelerate faster than controls.
- **Spending is up, but confidence is down**
  90% increased AI security budgets this year, yet 29% feel less secure than twelve months ago. The problem is outpacing the investment.
- **Most AI activity is invisible to security**
  94% of respondents report gaps in AI activity visibility. 88% cannot distinguish personal AI accounts from corporate instances. Only 6% claim to see the full scope of their organization's AI pipeline.
- **AI is rendering legacy data loss prevention powerless**
  DLP matches patterns while AI transforms meaning; only 8% have controls that evaluate content semantically, regardless of how it has been rewritten.
- **Agents act without guardrails**
  AI agents have write access to collaboration tools (53%), email (40%), code repositories (25%), and identity providers (8%). 91% of organizations only discover what an agent did after it has already executed the action.
- **Too much AI security runs on trust**
  31% rely on written policies and employee compliance as their primary enforcement. Another 11% have nothing at all. Only 23% say they enforce AI security inline, at the point of action.

AI-driven risk is expanding from human misuse to machine autonomy, and the controls are still working to address the first challenge. The survey points to four architectural priorities: continuous visibility into all AI activity including agent and M2M traffic, inline enforcement without creating friction and latency, semantic-aware data controls that evaluate meaning rather than patterns, and extending zero trust to non-human identities (NHIs). The chapters that follow measure how mature most organizations are, and how to close the execution gap.

AI Governance Lags Far Behind Adoption

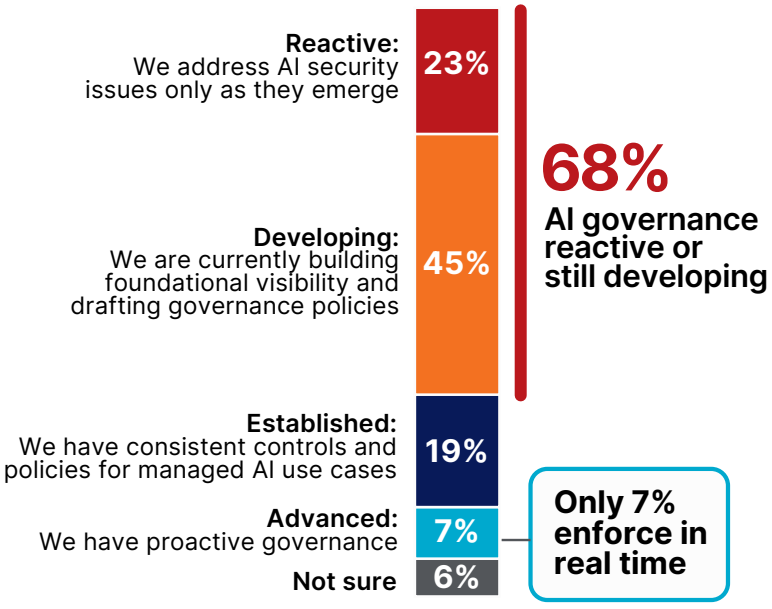
Twelve months ago, most organizations treated AI governance as a future priority: something to formalize once adoption stabilized. Adoption didn't wait. Copilots, code-completion tools, and content generators shipped into production across departments, and by the time security had a framework in place, the AI footprint was already operational.

Today, 68% of organizations describe their AI governance as reactive or still developing. Only 7% have reached advanced maturity with real-time policy enforcement. The 66-point gap between the 73% deploying AI tools and the 7% governing them in real time is a structural mismatch—organizations are building at production speed on a security and compliance foundation that barely exists. And the consequences are starting to show, with 39% having already experienced an AI-related near-miss involving unintended data exposure. Of those, 17% changed nothing afterward.

While the 68% of organizations with reactive governance or governance that is still in development may not sound alarming, it masks a problematic pattern inside many organizations. More than a third report fragmented AI adoption, with multiple teams deploying tools independently under no shared framework, standards, or security policies. One division runs autonomous agents under informal guidelines while another hasn't documented which AI tools employees use at all. The governance conversation isn't just behind schedule; in many organizations it never started. 48% predict that governance failures—specifically shadow AI and over-permissive access—will trigger the next major AI-related breach. The practitioners closest to the problem already know that one of their biggest risks comes from the tools their own employees adopted last quarter without telling anyone. That governance deficit cascades into every security layer that follows.

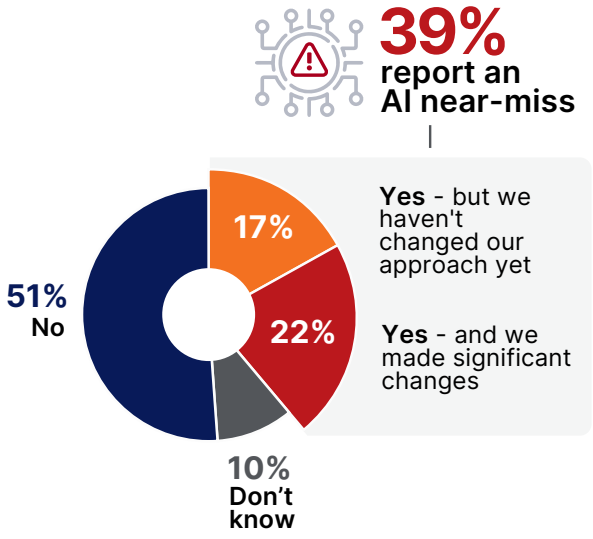
Real-time governance is rare

Which best describes your organization's maturity in governing and securing AI implementations and data?



Near-misses are already real

Have you experienced an AI-related "near miss" involving the unintended exposure or leakage of sensitive data that caused you to fundamentally reconsider your security approach?



The simplest structural fix: identify the three highest-risk AI use cases in your environment, embed enforceable policies for those three into technical controls, and assign an owner for each.

More Budget, Less Confidence

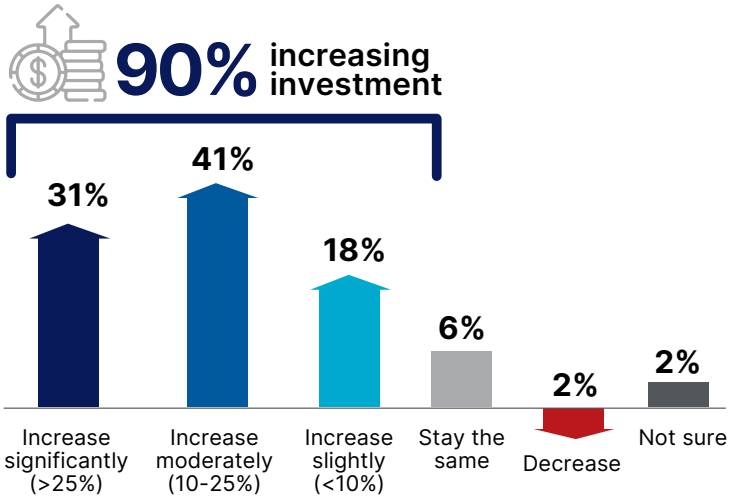
Paradoxically, the AI governance gap exists despite organizations investing more than ever in security. 90% increased AI security spending this year, with nearly a third raising budgets by more than 25%, yet 29% report feeling less secure than twelve months ago. Investment is increasing—confidence is not.

Research participants explained why: 34% see the biggest barrier coming from business pressure to adopt AI faster than security can follow. Skill gaps came second at 25%, and legacy tools that cannot interpret AI-specific threats ranked third at 21%. Budget challenges placed fourth at 14%. While the budget has arrived for many, the architecture it funds still reflects a pre-AI threat model.

Existing security tools were designed for known file formats, predictable data flows, and human-speed interactions. Adding more budget to that stack buys more of what already fails against AI-driven risk. The confidence breakdown confirms it: 51% rate their technical controls as weak, 50% rate visibility as weak, and 61% rate NHI governance as weak.

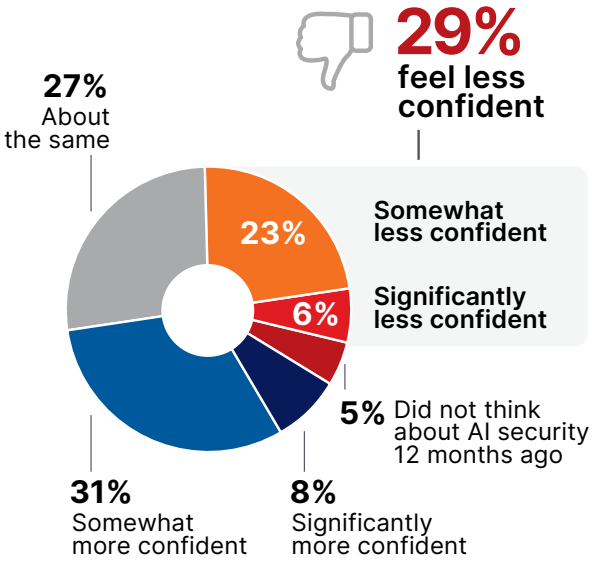
Spending is rising fast

How will your organization's investment in AI-specific security controls change over the next 12 months?



Confidence is slipping

Compared to 12 months ago, how has your confidence in your organization's ability to defend AI systems against attacks and data leakage changed?



Redirecting budget starts with mapping current AI security spend against those three weakness ratings (visibility, technical controls, and NHI governance) and concentrating investment where the gap between spending and capability is widest.

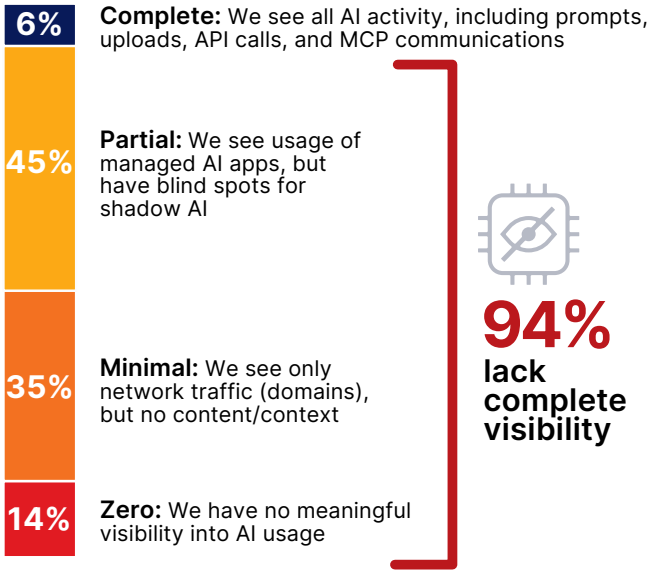
Most AI Activity Is Invisible to Security

You cannot secure what you cannot see. Only 6% of organizations report complete visibility into AI usage across their environment. 45% have partial visibility limited to managed applications, blind to anything outside of authorized tools. 35% see only network-level traffic patterns, enough to know something is happening but not what. The remaining 14% have no visibility at all. 94% are making AI security decisions with an incomplete picture, and for most, this is the default operating state.

Even where detection exists, distinguishing what matters remains difficult. 88% cannot reliably tell personal AI accounts from corporate instances on the same platform, the #1 technical blind spot in the survey. When a security team cannot tell whether an employee is using an authorized AI tenant or a personal account with no data governance, DLP policies, access controls, and audit trails all become unreliable. Shadow AI adds another layer to these blind spots: 31% rely on log review after the fact to identify unauthorized AI tools, 21% cannot detect shadow AI at all, and only 27% use a CASB or SWG for real-time discovery.

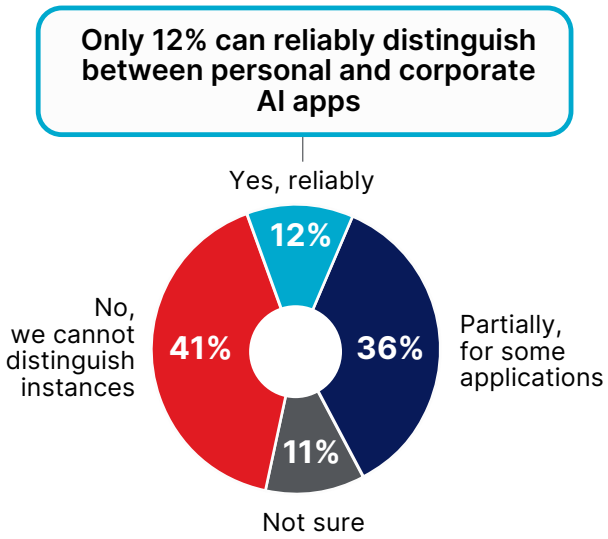
Visibility is mostly incomplete

How much visibility does your security team have into AI usage?



Instance identity is unclear

Can your security tools distinguish between personal and corporate instances of AI applications (e.g., personal ChatGPT vs. enterprise ChatGPT)?



Visibility is the structural prerequisite every other control depends upon. DLP, access policies, and acceptable use enforcement all assume the organization can see the activity it intends to govern. The hardest interactions to monitor are also the ones growing fastest: API integrations, MCP-based agent connections, and M2M communication dominate the difficulty rankings, while direct user-to-AI chat ranks last at 6%.

> Closing that visibility gap means extending activity-level monitoring to those channels, starting with account-level distinction between personal and corporate AI accounts as the foundation everything downstream depends on.

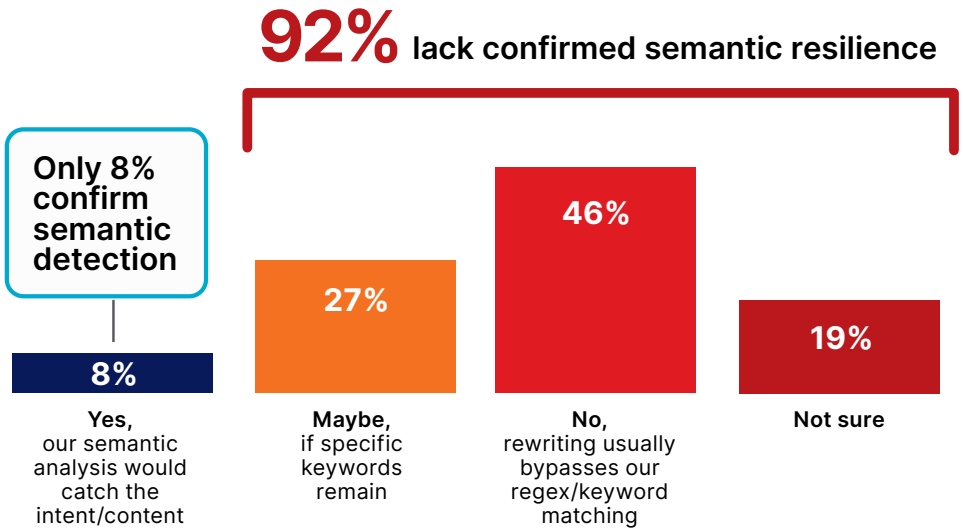
AI Renders Legacy DLP Powerless

Even where organizations can see AI activity, the primary tool tasked with catching data in motion was designed for a fundamentally different kind of movement. DLP was built to find specific patterns: credit card formats, Social Security number sequences, regex matches against known sensitive content. While DLP may block the upload or the copy/paste of sensitive data into prompts by looking for these patterns, if AI gets hold of the data, it will rephrase sensitive content—retaining its meaning—while discarding its original digital fingerprint.

The distinction is architectural. DLP operates at the syntactic layer, matching character sequences against predefined rules. AI operates at the semantic layer, transforming content while preserving intent. A simple transformation test makes it concrete: If an employee takes a secret project description and asks an AI to "summarize this into a professional email," the AI may replace "Project X" with "our upcoming strategic initiative." To a regex filter, "strategic initiative" looks perfectly safe, even though the semantic value (the secret) remains identical. Similar issues emerge when translating English secrets into a second language and back. AI generates a version of the data where every original keyword has been replaced, yet the underlying risk remains unchanged. Traditional DLP also fails at inference. It may scan two separate lists—one of names and one of medical conditions—and find them harmless in isolation, but an AI can rephrase the document to explicitly link the two, creating a HIPAA violation that a pattern-based DLP filter cannot see. 46% said their controls would miss these kinds of policy violations because rewriting typically bypasses regex and keyword matching. Another 27% said detection works only if specific keywords survive the transformation, a control that functions only when the adversary cooperates. Combined with the 19% unsure of their coverage, 92% of organizations lack DLP confirmed to work after AI rephrases content.

Rewriting defeats pattern controls

▶ The Transformation Test: If an employee asks an AI to "rewrite this document as a generic blog post," can your security controls detect the sensitive data in the output?



DLP still catches pattern-based violations, but AI-transformed content passes through undetected. The problem has moved to a layer DLP was never designed to inspect. The way to measure exposure is concrete: run the above transformation test against your own stack, take a classified document, ask an AI tool to rephrase it, and see whether your controls flag the output. That result becomes the baseline for deploying content-aware inspection that evaluates meaning at the point of transfer.

AI Agents Run Unsupervised

While data leakage through AI tools is the risk most organizations recognize, the deeper exposure is that AI systems are now acting on their own, with many operating in shadow mode outside security's view.

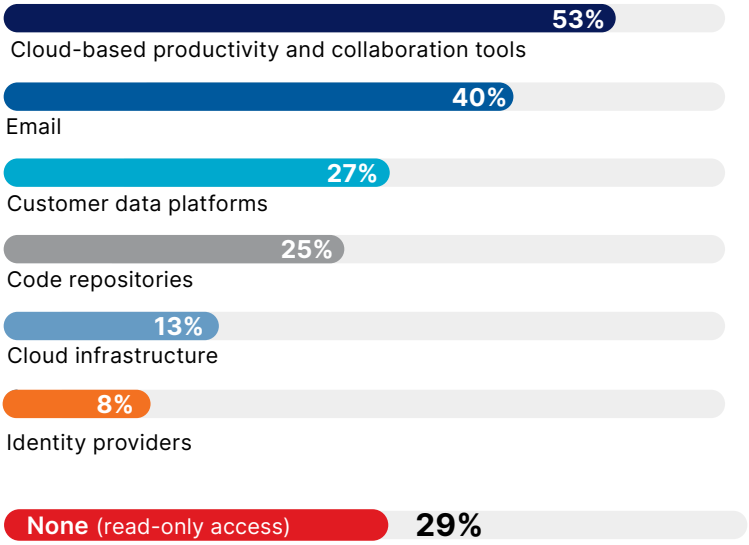
The survey quantifies how far this has spread. 56% report real agentic AI risk exposure: 24% in limited production, 9% at scale handling core business logic, and 23% through shadow deployments IT doesn't know about. 32% have zero visibility into agent actions, and 36% are blind to M2M AI traffic entirely.

Organizations that cannot see agents cannot know they have shadow agents. 10% say they have banned agentic AI, yet across the full survey population 23% report shadow use. In practice, bans often drive activity underground, making it harder to govern and even harder to contain when something goes wrong.

The write-access exposure is broader than most security teams expect. 53% grant AI tools write access to cloud productivity and collaboration suites, 40% to email, 25% to code repositories, and 13% to cloud infrastructure. Then there are the entries that change the nature of the risk: 8% grant write access to identity providers. An agent with write access to the identity layer can create service accounts, elevate privileges across federated systems, and grant themselves external access through API calls that never cross a network perimeter.

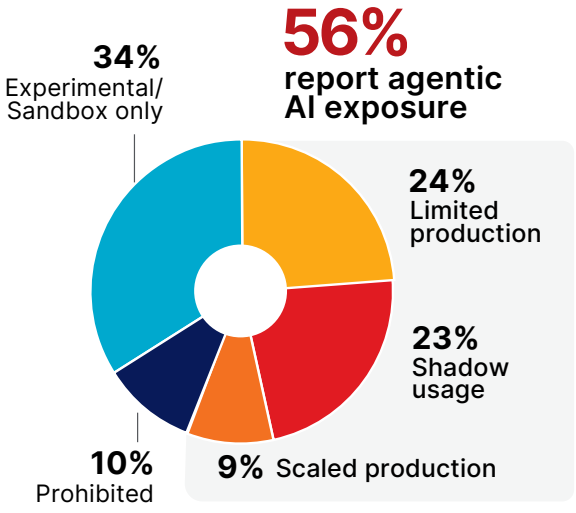
Agents already have write power

Which internal systems do your AI tools/agents have write access to?



Shadow deployment is common

How would you describe your adoption of "agentic AI" (AI that pursues goals independently)?



Only 29% of organizations limit AI tools to read-only access. For the remaining 71%, the remediation path is clear: audit which AI tools hold write access today and establish approval gates for any action that creates accounts, modifies permissions, or moves data externally.

When Agents Act, Nobody Stops Them

Agents have broad access across enterprise systems, and almost none of it can be intercepted. Once an agent initiates a harmful action, only 9% of organizations can intervene before it completes. The remaining 91% breaks down into degrees of helplessness: 24% can block some agent actions but not all, 35% would find the action only in logs after it completed, and 32% have no visibility into agent actions whatsoever. For every ten organizations running agentic AI, fewer than one can stop an agent from deleting a repository, modifying a customer record, or escalating a privilege before the action executes.

The consequences are already materializing. 37% experienced AI agent-caused operational issues in the past twelve months, with 8% significant enough to cause outages or data corruption. 38% point to an agent autonomously moving data to an untrusted location as their top runaway concern, and 24% fear an agent deleting critical configurations or code. These concerns are reflected in independently reported 2025–2026 events. In mid-2025, the EchoLeak vulnerability (CVE-2025-32711, CVSS 9.3) demonstrated a zero-click prompt injection against Microsoft 365 Copilot, enabling enterprise data exfiltration without user interaction. In early 2026, researchers disclosed the Reprompt attack, which chained three techniques to turn Copilot Personal into a single-click data exfiltration channel.

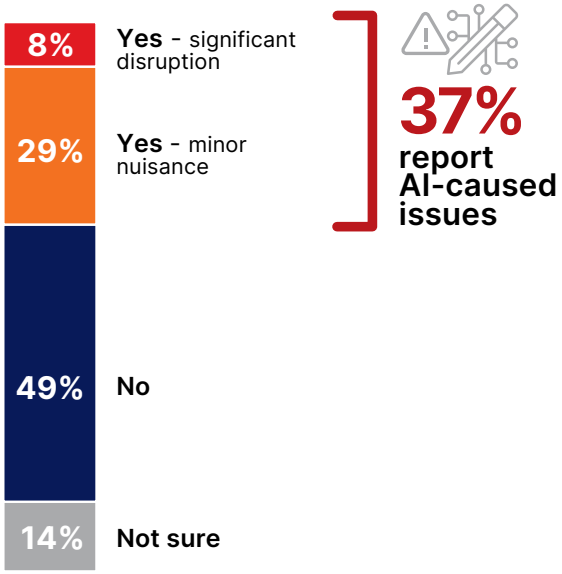
Somewhere in that cohort lacking AI agent visibility (32%) sits a SOC analyst who will arrive Monday morning, trace an anomalous privilege change to a service account created by an agent 72 hours earlier, and discover the agent has been writing to production systems all weekend. The logs will show every action. No alert fired because no detection rule existed for agent-initiated behavior.

Prevention is the exception

Can your organization prevent a risky AI-driven action (e.g., an agent deleting a repo) before it

Operational fallout is showing up

In the past year, has an AI tool caused an operational issue?



That gap has a fix: define what anomalous looks like for agent actions in your environment, build detection rules for those patterns, and require human-in-the-loop approval for high-risk agent actions like account creation, permission changes, and external data transfers. Automated interception at the request layer is the target state as tooling matures.

Zero Trust Stops at the Machine

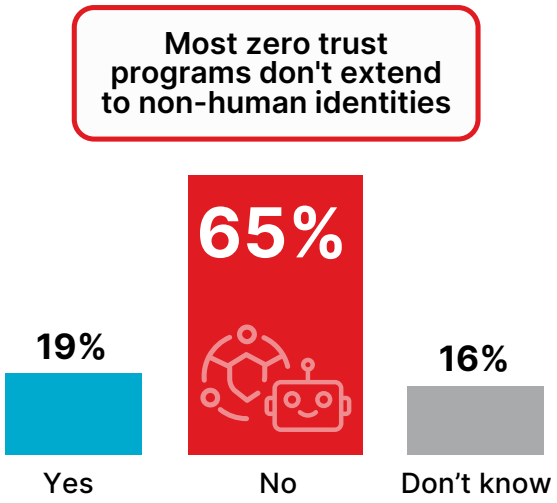
91% of organizations cannot stop an agent before it acts. The reason is architectural: zero trust was built around a user with a device, a location, a behavior pattern, and a risk score. An AI agent has a credential, a scope, and a task. 62% apply zero trust principles to AI security in some form, making it the most widely adopted approach. Yet 65% say their current zero trust controls cannot secure non-human identities (NHI).

NHI governance scores lowest across all dimensions measured, with 61% rating it weak, yet 78% expect NHI growth to outpace human identity growth over the coming year. Every new agent, microservice, and automation workflow creates service accounts and API keys that traditional identity governance was never designed to handle. These frameworks were built for entities that persist across business quarters. An agent identity may persist for minutes.

The protocols agents use to communicate create a second gap. MCP has emerged as a common connector between AI agents and enterprise tools. In many current implementations, interoperability has taken priority over built-in identity verification, least-privilege enforcement, or independent audit visibility. The survey shows that only 8% of organizations have policies governing MCP. The remaining 92% are either not monitoring it or have never heard of it. 36% have no visibility into M2M AI traffic at all, and another 28% rely entirely on vendor platform security without independent verification. Only 14% inspect API traffic with the same rigor applied to user traffic.

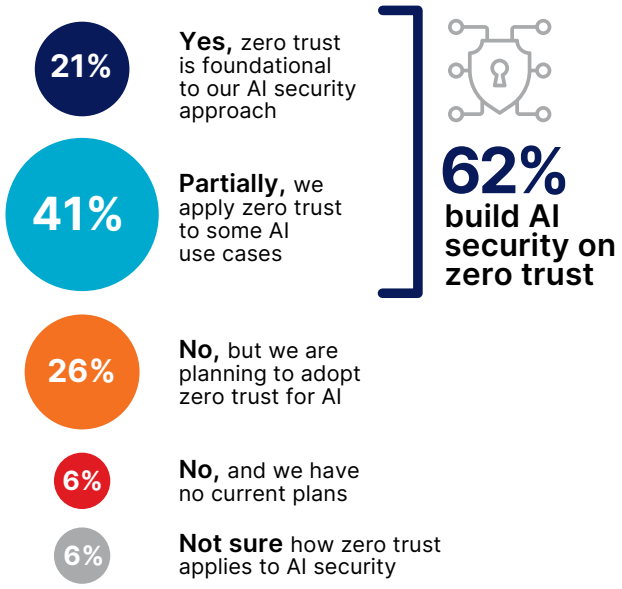
Non-human identity is out of coverage

Do your current zero trust access controls enable your organization to secure non-human identities?



Zero trust is still the strategy

Is your AI security strategy built on zero trust principles (verify every request, monitor every data flow, grant access based on dynamic risk)?



Addressing this gap requires an organization to converge the protocol layer and the identity layer so that agent credentials, scopes, and permissions are evaluated with the same rigor applied to human identities.

Most AI Security Runs on Trust

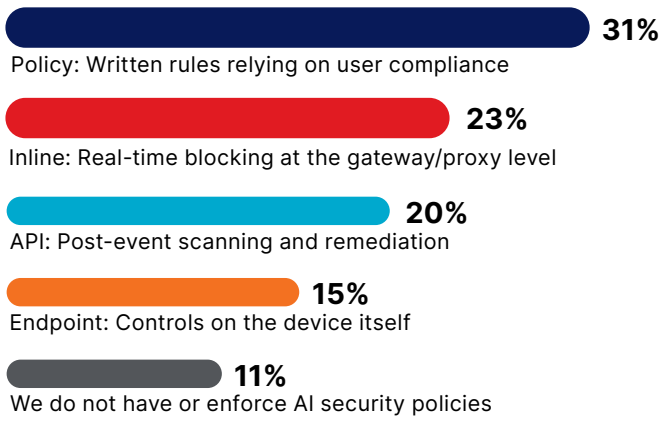
Zero trust covers human identities but leaves AI agents and NHI largely ungoverned. The question for this gap is practical: when an AI tool violates a policy or an agent takes a risky action, what enforcement mechanism actually intervenes? The survey maps the full enforcement spectrum. 31% enforce AI security through written policies and employee compliance. Another 20% rely on post-event API scanning, catching violations after the action completes. Endpoint-based controls account for 15%, and inline real-time enforcement for 23%. The remaining 11% have no AI security policies at all. The largest single enforcement category is the honor system. The second-largest is reviewing what already happened, after the fact.

Closer scrutiny of the data suggests that even the 23% of organizations reporting inline enforcement may be operating blunt instruments. 42% control AI applications through binary block-or-allow for entire platforms, with no way to allow only corporate accounts while blocking personal accounts, or to allow a research query while blocking the upload of a financial model. Only 19% have granular, activity-level controls that distinguish actions within an allowed application. Where real-time controls exist, they focus on human actions: file upload blocking (48%) and paste detection (37%) lead, while content posting (29%) and download controls (25%) trail behind. Agent-initiated API calls, OAuth token exchanges, and M2M data flows pass through largely unaddressed.

This execution gap is a coherence failure. When the CASB, the DLP engine, and the access policy each see a fragment of the picture, none can enforce the full intent. A simple diagnostic test: if a single policy decision requires data from more than two consoles, that fragmentation is where enforcement breaks down. Closing that gap means unifying those layers so a single evaluation draws on content classification, user identity, and AI instance type before an action completes.

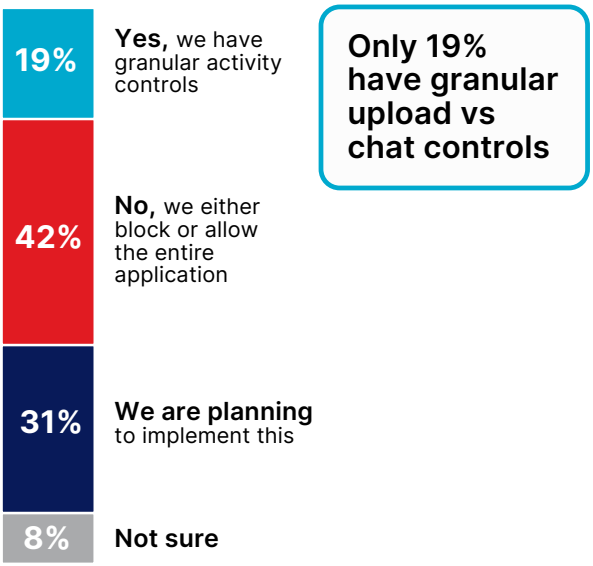
Enforcement is mostly post-action

How are your AI security policies primarily enforced?



Controls are still blunt

Do you enforce different policies for "Upload" vs. "Chat"?



AI security cannot be layered onto legacy perimeter models. It requires cloud-native inline enforcement that evaluates identity, content, and context in a single decision point before execution occurs.

How Mature Is Your AI Security?

Each gap examined so far amplifies the others: weak visibility undermines DLP, ungoverned agents bypass access controls, and fragmented enforcement leaves every layer exposed. The Capability Maturity Model below maps six core AI security domains across three tiers of readiness. Each cell describes the capabilities at that stage. Find the description that matches your organization in each row. The domain with the lowest maturity is your weakest link and most likely point of failure. That is where investment should go first.

AI SECURITY DOMAINS	REACTIVE	MANAGED	ADAPTIVE
Governance & Risk Alignment	Policies on paper. Enforcement inconsistent.	Policies enforced for managed AI tools. Shadow AI ungoverned.	Policy embedded in technical controls, real-time.
Visibility & Situational Awareness	Partial or no visibility. Cannot distinguish personal from corporate instances.	Activity-level awareness across managed SaaS and APIs. Personal and corporate instances distinguished.	Real-time visibility across all AI workflows including agents and machine-to-machine.
Data & Asset Protection	Pattern-based controls fail under AI transformation.	DLP extended to AI traffic including chat and prompts. Semantic detection in pilot.	Semantic inspection across all AI data flows.
Access & Execution Control	Post-execution enforcement. Honor system for policy enforcement.	Inline enforcement for human-initiated AI. Agent actions logged, not blocked.	Dynamic pre-execution enforcement, human and non-human.
Detection & Response	Log-based monitoring. No detection logic specific to AI-driven behavior.	Detection rules for known AI misuse patterns. Manual containment.	Continuous monitoring with automated containment across all AI activity.
Architectural Integration & Operational Resilience	Fragmented. Visibility, protection, enforcement in silos.	Controls integrated for managed SaaS. Gaps in API and agent traffic.	Unified framework durable under automation and scale.

When asked about their regrets regarding AI adoption 38% wish their governance had preceded adoption of AI at scale, and 25% wish they had invested in visibility controls sooner. Only 7% report satisfaction with their current approach—the lowest confidence indicator across the entire survey.

Change is driven by pressure

52%
say regulation forces change

47%
say a breach forces change

Closing the Execution Gap

The Capability Maturity Model shows where the gaps are. The following summarizes the most useful actions to improve across each risk vector, starting with visibility—the foundation every other control depends on.

1. **Close AI Visibility Gaps:** 94% report gaps; 88% cannot distinguish personal from corporate accounts.
   Expand activity-level monitoring across SaaS, API, and M2M traffic, starting with account-level distinction between personal and corporate AI accounts: the prerequisite for reliable DLP, access controls, and audit trails.
2. **Translate Policy Into Enforceable Guardrails:** 68% govern reactively; only 7% enforce in real time.
   Identify the three highest-risk AI use cases in your environment, embed enforceable policies for those into technical controls, and assign an owner for each before expanding coverage to all remaining AI use cases.
3. **Deploy Semantic Data Protection:** 46% fail the content transformation test.
   Run the content transformation test against your own DLP: take a classified document, ask an AI tool to rephrase it, and see whether your controls flag the output. That result is the baseline for deploying content-aware inspection that evaluates meaning at the point of transfer.
4. **Enforce Before Execution:** 23% enforce inline; 9% can stop a risky agent action in advance.
   Audit which agents have write access today and establish approval gates for any action that creates accounts, modifies permissions, or moves data externally.
5. **Modernize Detection and Containment:** 67% rely on logs or have no visibility into agent actions; 37% already experienced AI-caused operational issues.
   Define what anomalous looks like for agent actions in your environment and build detection rules for those patterns. Establish containment playbooks that can intercept at the request layer before an action completes, rather than generating a ticket after the damage is done.
6. **Reduce Control Fragmentation:** 42% use binary block-or-allow controls with no activity-level differentiation; only 14% inspect API traffic with the same rigor applied to user traffic.
   Unify CASB, DLP, and access policy so a single evaluation draws on content classification, user identity, and AI instance type. If a policy decision currently requires data from more than two consoles, that fragmentation is where enforcement breaks down.

AI security is an operational discipline now. The maturity dimensions are mapped, the dependency sequence is clear, and the actions are concrete. What remains is the decision to build.

Governance is behind

68%
are reacting to or developing governance and security policies for AI implementations and data

Visibility is incomplete

94%
lack complete visibility into AI usage

Data controls don't survive rewriting

Only **8%** confirm semantic detection of intent/content

Interception is rare

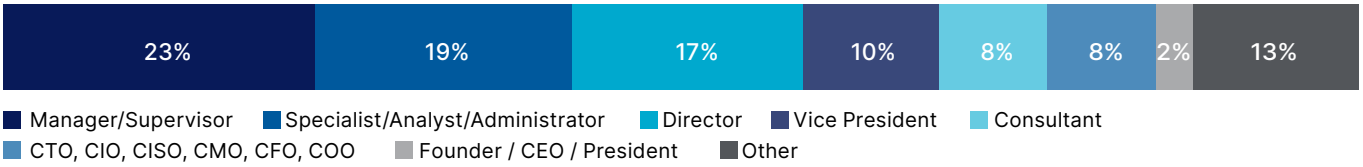
Only **9%** have inline prevention for API/agent actions

Methodology and Demographics

This report is based on a survey of 1,253 cybersecurity and IT professionals conducted in early 2026. Respondents represent security practitioners, architects, and technology leaders responsible for protecting enterprise infrastructure, cloud environments, and AI-driven applications across a wide range of industries and organization sizes.

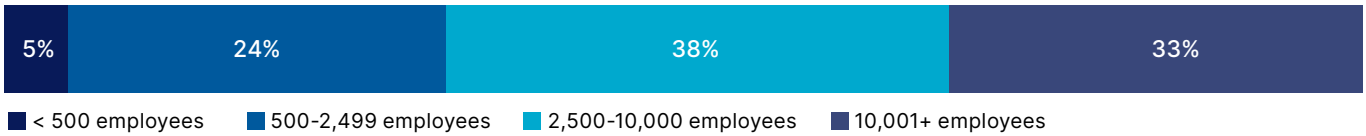
The research examines how organizations are securing AI deployments, focusing on governance maturity, visibility into AI activity, data protection, non-human identity management, and the control of autonomous agents. Using a stratified sampling approach, the survey achieved a 95% confidence level with a margin of error of +/- 2.8%.

CAREER LEVEL

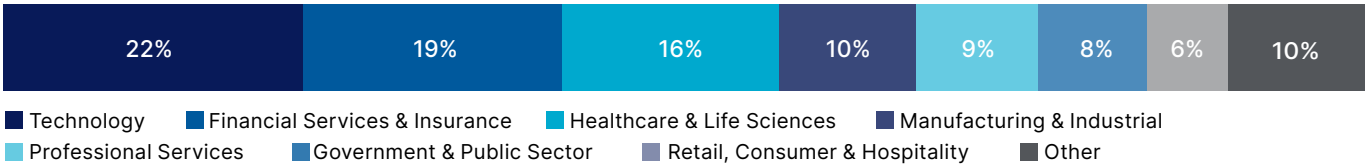


DEPARTMENT

COMPANY SIZE



INDUSTRY



About Netskope

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.

To learn more visit

netskope.com/ai

Cybersecurity
INSIDERS

BENCHMARK YOUR SECURITY MATURITY

Independent cybersecurity research revealing the gaps that shape cybersecurity strategy

Cybersecurity Insiders produces independent research based on surveys of cybersecurity leaders and practitioners worldwide. Our reports reveal where security strategies break down in practice — helping organizations benchmark their maturity, identify capability gaps, and prioritize the actions needed to close them.

For more information, visit

cybersecurity-insiders.com

April 2026 // Version 1.7