

ADOPTION IS OUTPACING CONTROL.

The State of AI Agent Security 2026

AI agent adoption has accelerated, but security models have not evolved at the same pace. Our survey of 919 executives and practitioners reveals the structural gaps in identity, authorization, and runtime governance as AI agents move into production.

EXECUTIVE SUMMARY: THE MISMATCH THAT DEFINES AGENT SECURITY TODAY

AI agents are already embedded in production systems, interacting with APIs, tools, and other agents. While adoption has accelerated, security models have not evolved at the same pace, not because teams don't understand the risk, but because existing identity and authorization frameworks were not built for autonomous, agentic systems.

Adoption Outpaces Governance

81%

of teams are past the planning phase, yet only 14.4% have full security approval.

Incidents are the Norm

88%

of organizations confirmed or suspected security incidents this year.

The Identity Crisis

22%

Only 22% of teams treat agents as independent identities (most still rely on shared API keys).

This report synthesizes data from two distinct perspectives:

- **The strategic view** from an executive survey focused on deployment velocity, high-level governance, and organizational risk.
- **The practitioner view** from a hands-on technical survey of engineers and architects focused on identity, access control, and real runtime incidents

Together, they tell a clear story:

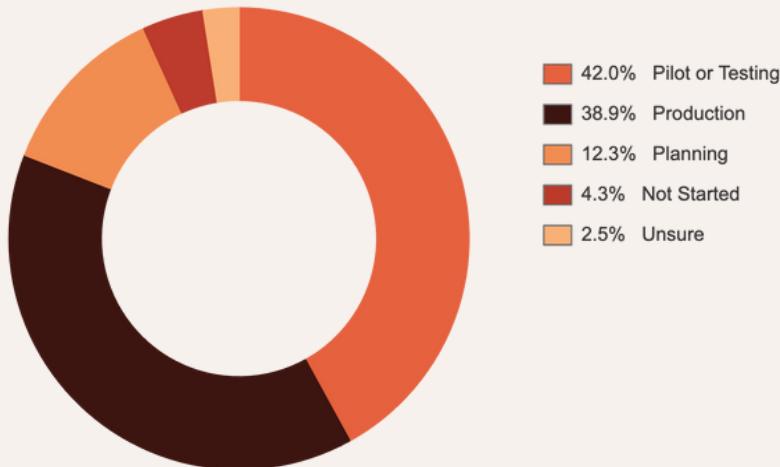
AI agent security is no longer a theoretical concern, and today's gaps are structural, not accidental.

AI Agents Have Quietly Become Production Infrastructure

AI agents are already deployed at meaningful scale

AI agents are no longer just experiments, they have become core components of distributed systems, behaving as autonomous infrastructure that inherits the same security expectations as any production service. The survey data confirms that 80.9% of technical teams have moved past the planning phase and are now actively testing or running agents in live environments.

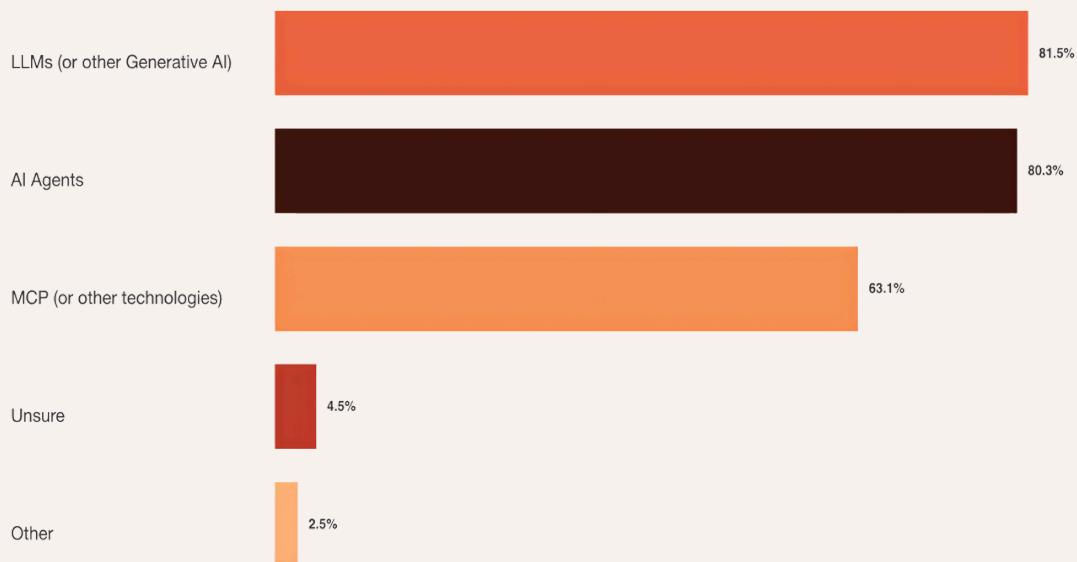
Technical survey – AI Adoption Journey



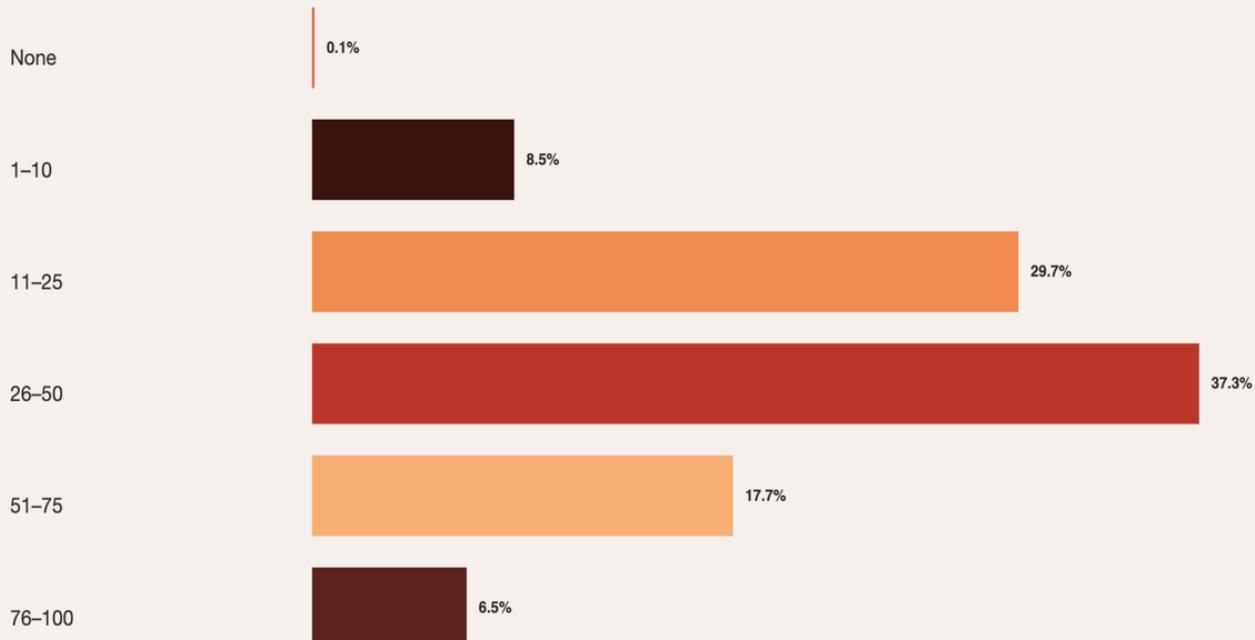
The shift toward agentic systems is driven by a move beyond using simply Large Language Models (LLMs). While the use of Generative AI is already mainstream, 80.3% are now specifically deploying AI Agents. This combined with the rapid adoption of the Model Context Protocol (MCP), indicates that the focus has shifted toward how agents connect to and interact with external tools and data.

The diversity of technologies used translates directly into volume. Organizations are not just managing a single "helper" agent, they are overseeing complex agent fleets. Our survey shows that **the average organization now manages 37 agents**.

Technical survey – AI Adoption Journey



Exec survey – Number of agents deployed



Beyond the Data

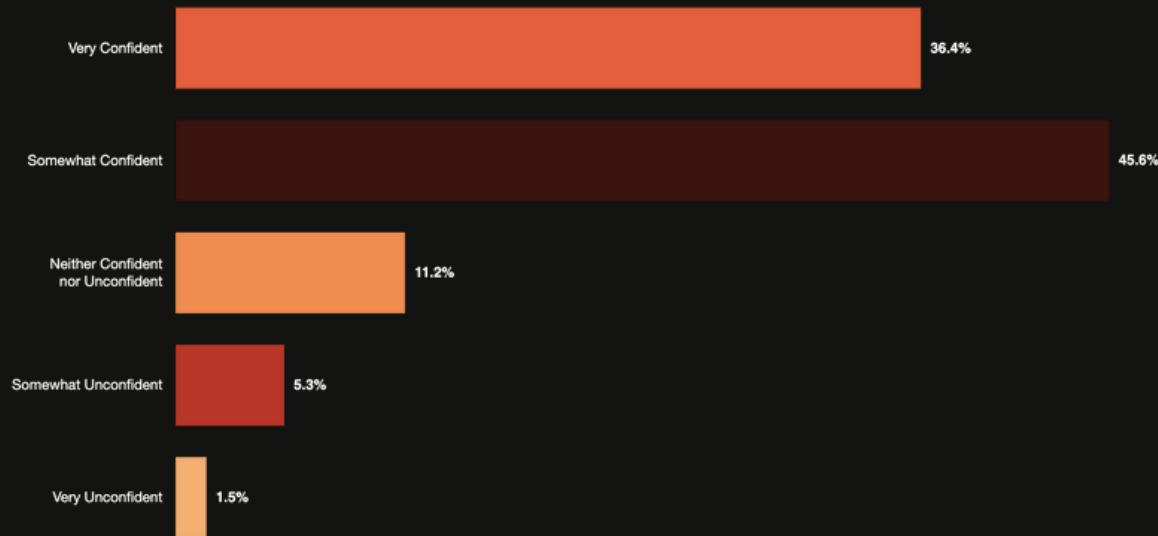
For a deeper dive into the future of autonomous systems and infrastructure, explore our [A2A Summit Hub](#), which features industry leaders discussing the next era of agent-to-agent communication.

Confidence Is High, but Coverage Is Partial

 **Most teams feel confident, even when half their agents are unsecured**

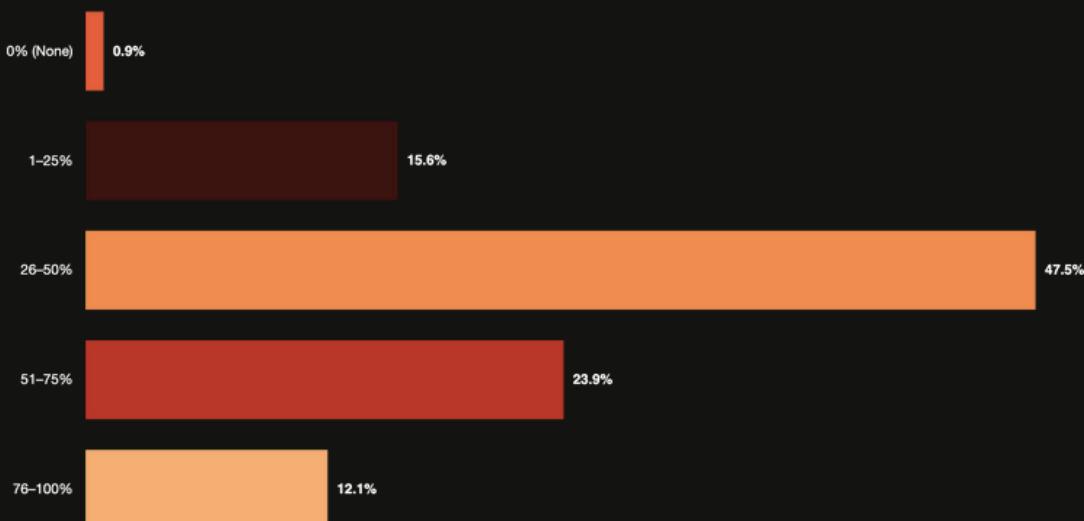
There is a dangerous disconnect between how secure organizations feel and the actual technical controls they have in place. 82.0% of exec respondents feel confident that their policies can protect against misuse or unauthorized agent actions. However, this confidence is often based on high-level policy documentation rather than real-time, granular enforcement at the API or identity layer.

Confidence in AI Security Policies



The high level of confidence begins to break down when we look at actual monitoring coverage. On average, only 47.1% of an organization's AI agents are actively monitored or secured. This means that more than half of AI agents operate without any security oversight or logging.

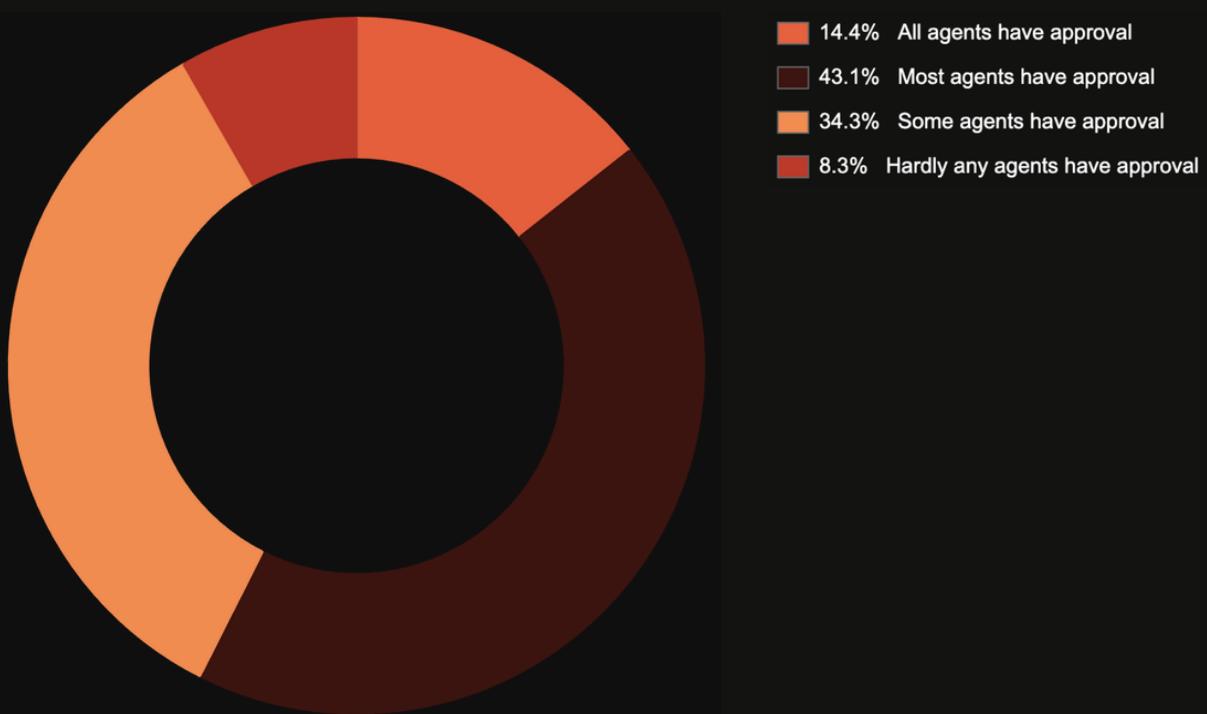
Percentage of AI agents actively monitored and secured



The Rise of Shadow AI: Approval Lagging Behind Deployment

This coverage gap is driven by a lack of centralized governance during the deployment phase. Our data shows that only 14.4% of organizations have achieved full IT and security approval for their entire agent fleet. The majority of agents are being deployed at the departmental or team level (often bypassing official security vetting entirely). This "Shadow AI" creates a scenario where agents are interacting with production data before the security team even knows they exist.

AI agents deployed with full approval from IT or security teams



Beyond the Data

To learn how to bridge the gap between high-level policy and technical enforcement, watch this webinar on [Securing AI Agents, Managing Identity, and Trust](#) for actionable strategies on building a trusted agentic ecosystem.

Incidents Are Already the Norm, Not the Exception

Security failures are no longer a theoretical risk, they are a widespread reality. An overwhelming 88% of organizations report either confirmed or suspected AI agent security or privacy incidents within the last year. If we look at respondents from the healthcare sector, the incident rate is even more alarming with a staggering 92.7% of healthcare organizations reporting or suspecting an AI agent security incident. This reflects the complexity of securing agents that interact with sensitive healthcare data.

59%

Confirmed Incidents

29%

Suspected Incidents

12%

No Reported Incidents

Practitioner Stories

The Over-Privileged Optimizer

"During a production rollout, we discovered that the AI agent that was supposed to only have read-only privileges was making API calls with elevated privileges beyond what was intended. This occurred because the agent's learning model dynamically adjusted workflows and attempted to optimize remediation speed by invoking administrative functions that were not part of its original scope."

VP, Director, or Manager | Financial Services | +10,000 employees.

The Sanitization Bypass

"We found a prompt injection vulnerability where user-supplied instructions bypassed our input sanitization layer and were forwarded directly to agent-to-agent communication channels, temporarily granting one agent unauthorized write access to user databases before our audit trail and circuit breaker mechanisms detected and halted the breach within 2 seconds."

C-Suite | Healthcare & Life Sciences | 1-100 employees

The Exfiltration Attempt

"We had built an agent to automate some tasks. The agent was connected to internal tools. As per instructions it attached some sensitive information and was trying to send outside to the organization. Luckily, another system blocked it and we caught it."

VP, Director, or Manager | Telecom, Media & Technology | 1,000-10,000 employees.

The Permission Leak

"The agent had broader permissions than necessary and was able to access internal test data beyond its intended scope. While no sensitive or customer-facing data was exposed, the behavior raised concerns around over-privileged access and insufficient guardrails. We identified the issue through internal monitoring and log reviews after noticing unexpected API calls."

Developer, Architect, or Engineer | Telecom, Media & Technology | 1,000-10,000 employees.

The Silent Scope Creep

"One security issue we encountered occurred during the early production rollout of an internal AI agent that had access to multiple backend services via API keys. ... we discovered that the agent was granted broader permissions than necessary due to a shared service account configuration. This meant that under certain prompt conditions, the agent could access endpoints outside its intended scope."

Developer, Architect, or Engineer | Manufacturing, Industrial & Transportation | 1500-1,000 employees.

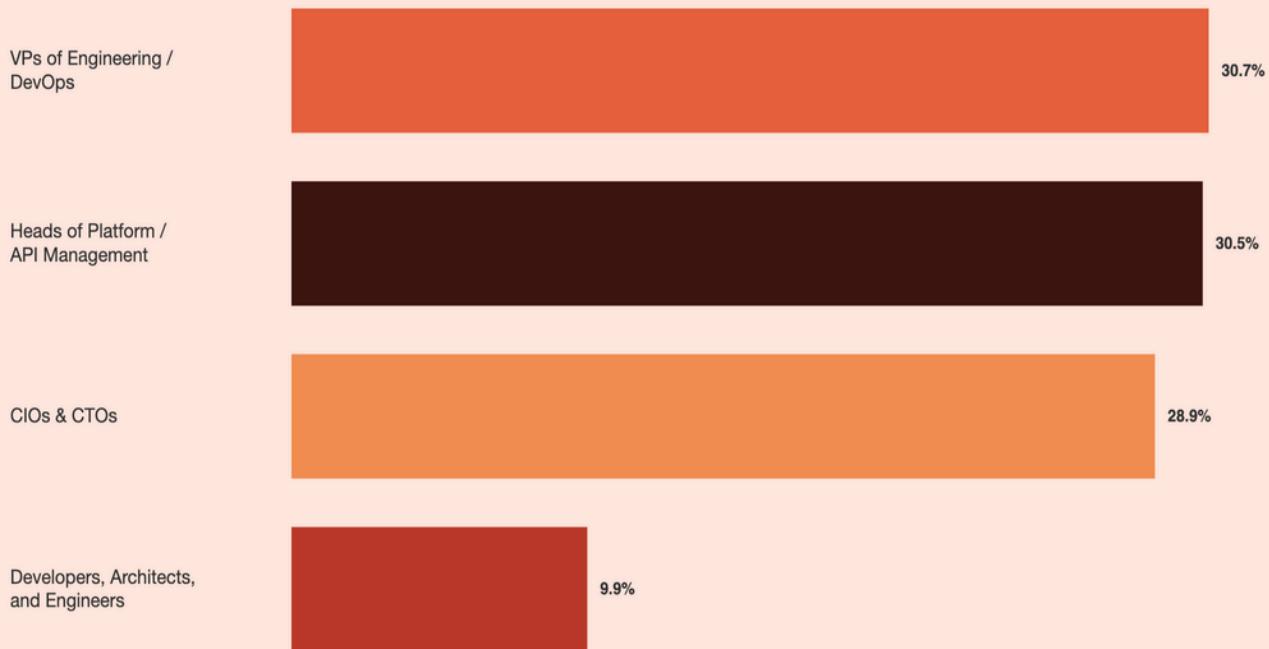
Beyond the Data

For a comprehensive breakdown of the vulnerabilities cited by the respondents, read this practical review of [OWASP Top 10 for Agentic Applications](#) to learn how to defend against these emerging threats

Survey Demographics: Representing the Enterprise AI Lifecycle

The insights in this report are derived from a survey of 919 participants, representing a deliberate balance between strategic leadership and the technical architects responsible for agentic infrastructure. This cross-sectional approach ensures the data reflects both organizational risks and ground-level execution realities.

Representing the Full Decision-Making Chain



Industry and Organizational Scale

The participation was distributed across a wide array of high-stakes industries, including **Telecommunications (23.6%)**, **Financial Services (20.8%)**, **Manufacturing (17.7%)**, **Healthcare (17.4%)**, and **Transportation & Logistics (16.3%)**. The survey also reflects a balanced representation of company sizes: while **40.4%** of respondents represent mid-sized organizations (**250–1,000 employees**), the remaining **59.6%** are larger enterprises (including 29.5% with 2,500–10,000 employees and 2.9% with over 10,000).

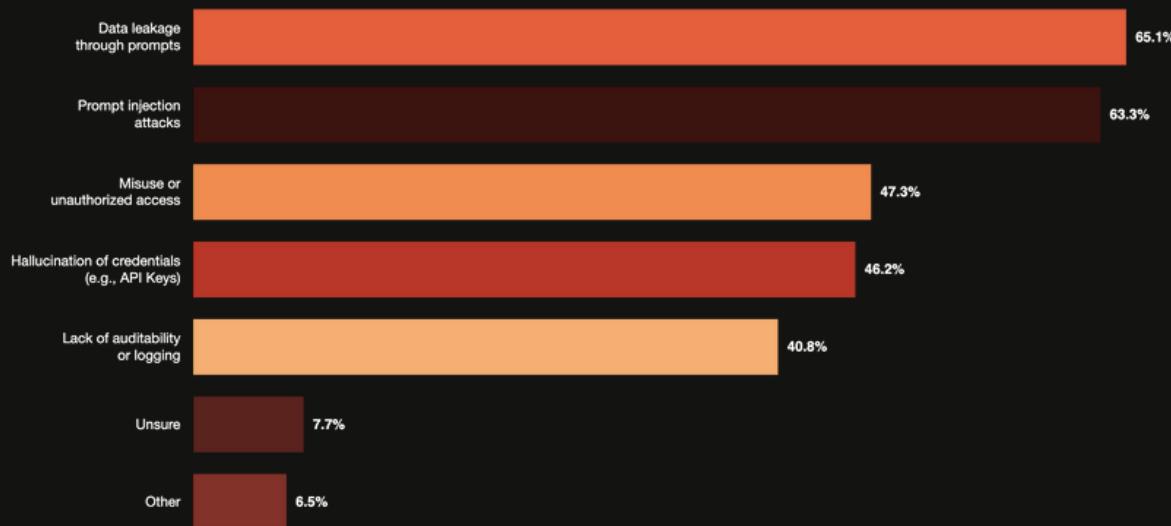
The Real Threat Model Is About Control, Not Model Quality

💡 Teams are worried about misuse, not hallucinations

Early discussions around AI security focused on "hallucinations" or model inaccuracies, but organizations moving agents into production are now prioritizing structural control. The primary risk is no longer that an agent might be incorrect, but that it is too efficient at performing actions it was never intended to do.

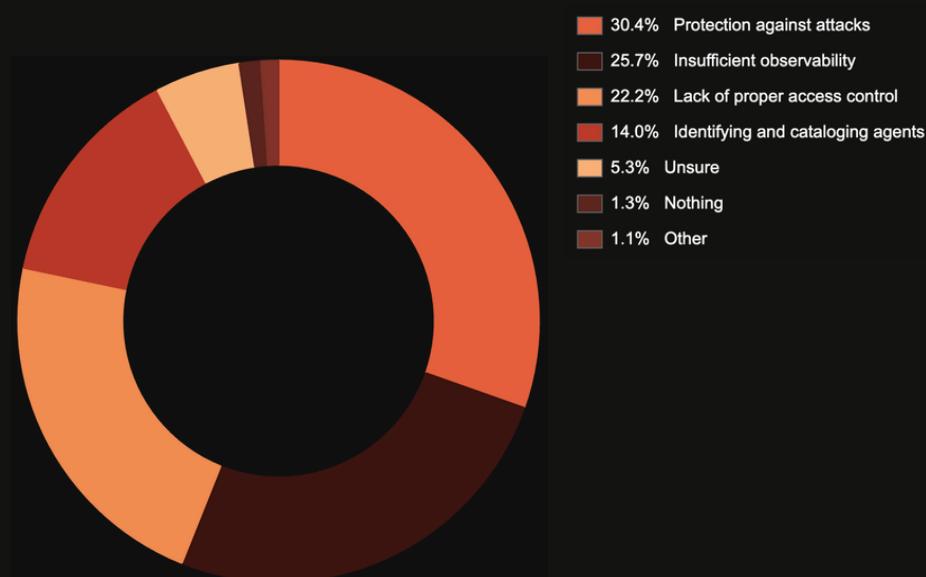
Technical teams identify data leakage and prompt injection as the most critical threats. Notably, nearly half of respondents now consider "misuse or unauthorized access of LLMs" as a top-tier concern, signaling a move toward traditional cybersecurity priorities.

4.1 Top AI Security Risks When Using LLMs



When building **AI Agents** and **MCP servers**, the focus shifts overwhelmingly toward **observability** and **access control**. Over half of builders (57.4%) cite a lack of logging and audit trails as a primary obstacle, highlighting a massive visibility gap in current agentic architectures.

Main Security Concerns When Building Agents & MCP Servers



Takeaway

The dominant risk is loss of control: who can do what, with which tools, and on whose behalf. This reframes agent security as an identity and governance problem, not an AI accuracy problem.

“

"They are thinking for you, they are taking the decision for you of what the right tool to call or other agent to call to complete the task. That's the whole point. They are doing the job."

Darrell Miller, Partner API Architect at Microsoft

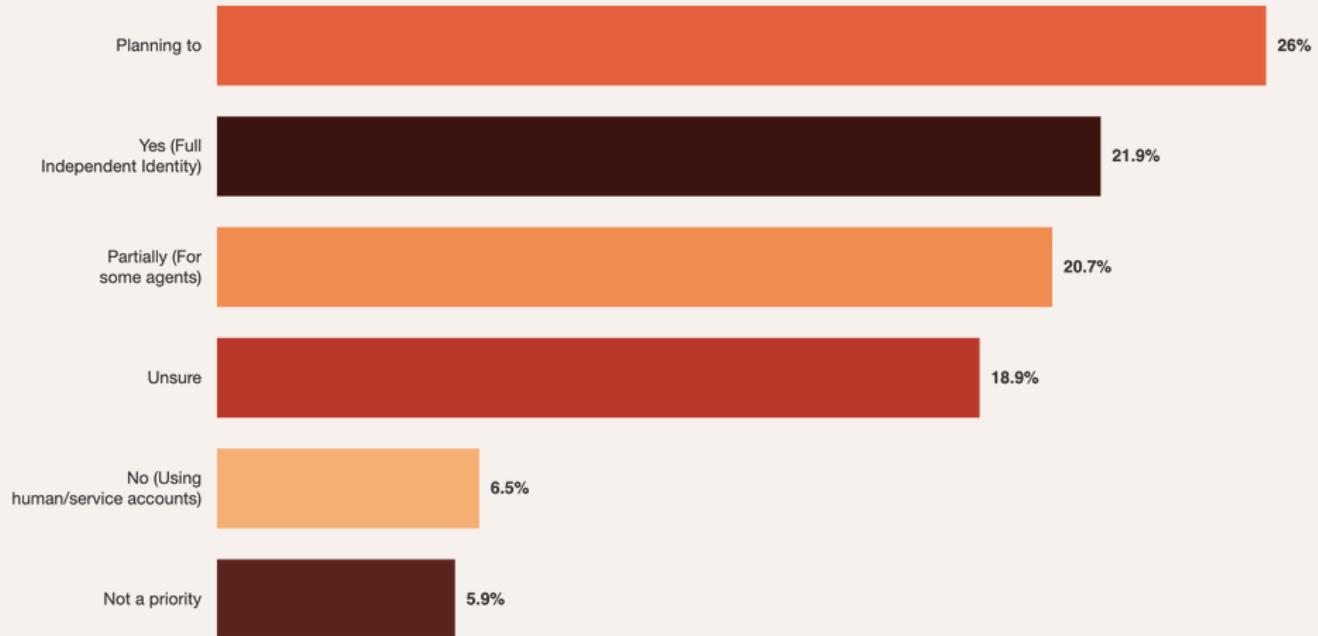
From APIs to Agents: The New Language of Enterprise Collaboration, [A2A Summit](#)

Identity Is The Weakest Link

 **AI agents are already deployed at meaningful scale**

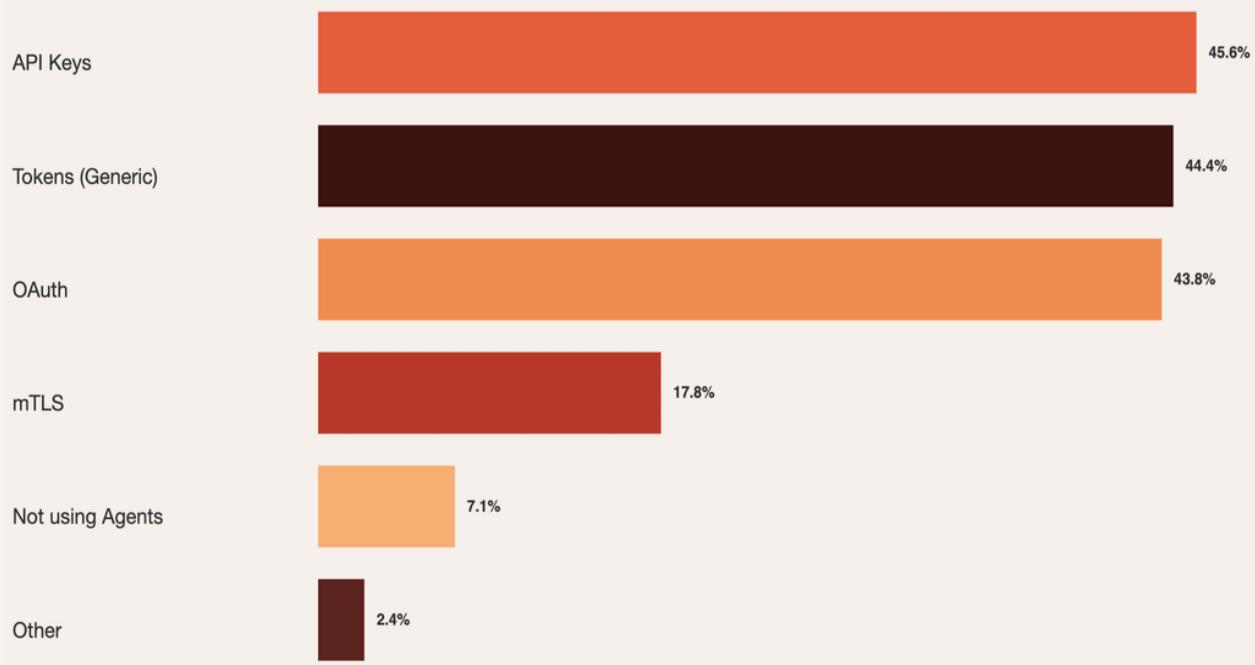
AI agents are rapidly becoming active participants in enterprise ecosystems, yet the foundational security principle of **unique identity** is largely being ignored. Only **21.9%** of respondents currently treat AI agents as independent, identity-bearing entities within their security model. Most organizations still treat agents as extensions of human users or generic service accounts, creating significant gaps in auditability and granular access control.

Treatment of AI Agents as independent, identity-bearing entities within the security model



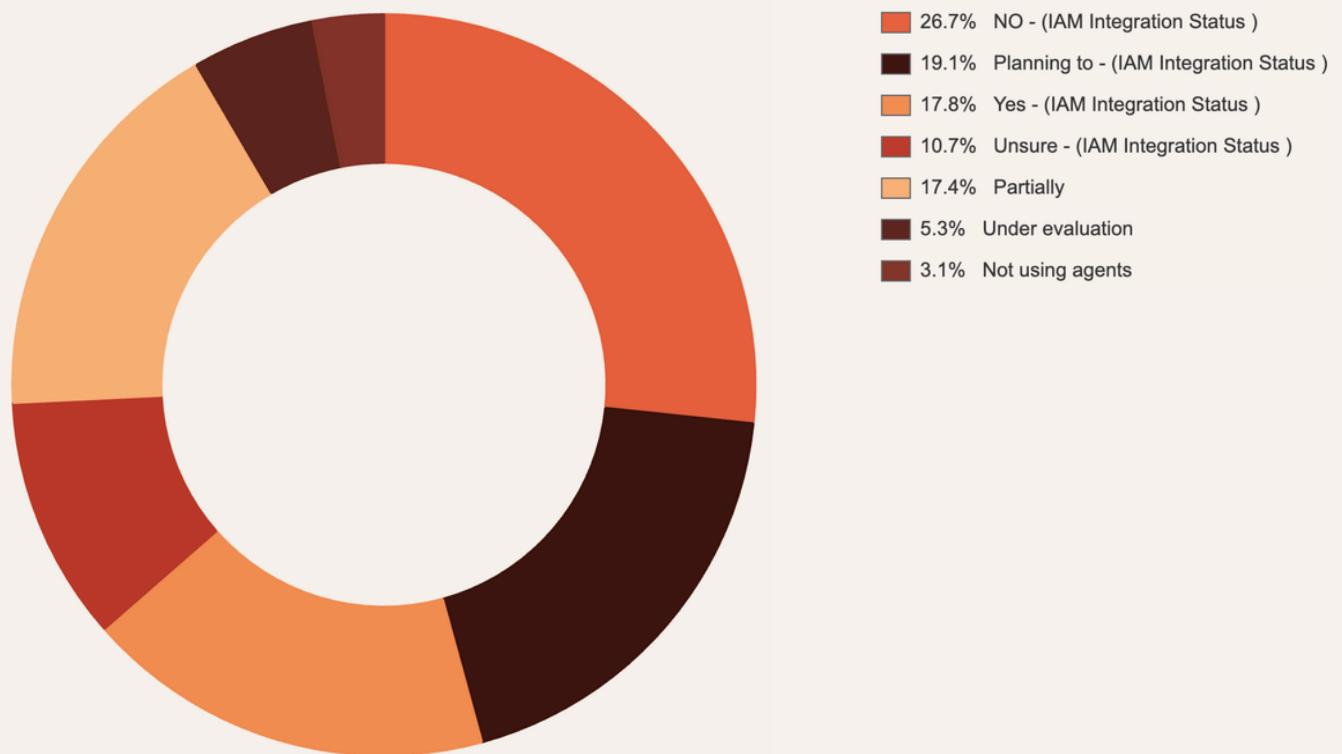
For agent-to-agent interactions, teams rely heavily on insecure or shared methods for authentication like **API Keys (45.6%)** and **Generic Tokens (44.4%)**, while secure standards like mTLS are utilized by only **17.8%**.

Agent-to-Agent Authentication Methods



Integration with existing corporate identity systems is lagging, with only **23.7%** of organizations using their existing **IAM/IdP** as an authorization server for their agentic (MCP) infrastructure.

Use of IAM/IdP as Authorization Server for MCP Servers



Real-World Evidence

Practitioner stories

"Honestly, general LLM security is still a concern on an enterprise level so we have all been using our own personal accounts with the agents. Therefore, we haven't yet given much focus to agent security since we are still finalizing our workflows. Appears to be a gap in internal knowledge on this subject."

VP, Director, or Manager | Telecom, Media & Technology | 1,000-10,000 employees

Practitioner stories

"During testing, we discovered that the agent was granted broader permissions than necessary due to a shared service account configuration. This meant that under certain prompt conditions, the agent could access endpoints outside its intended scope... This experience reinforced the need to treat AI agents as first-class security principals."

Principal or Lead | Financial Services | 1,000-10,000 employees

Practitioner stories

"We noticed that some agents were sharing passwords to access internal tools, which is a security risk. We addressed it by disabling shared accounts, creating individual logins, and conducting training on secure access practices."

VP, Director, or Manager | Financial Services | 10,000+ employees

Beyond the Data

Without strong, explicit agent identities, delegation becomes opaque, accountability breaks down, and audits lose meaning. Visit our [Agentic IAM Learning Hub](#) to learn how to treat agents as first-class, identity-bearing entities within your security model.

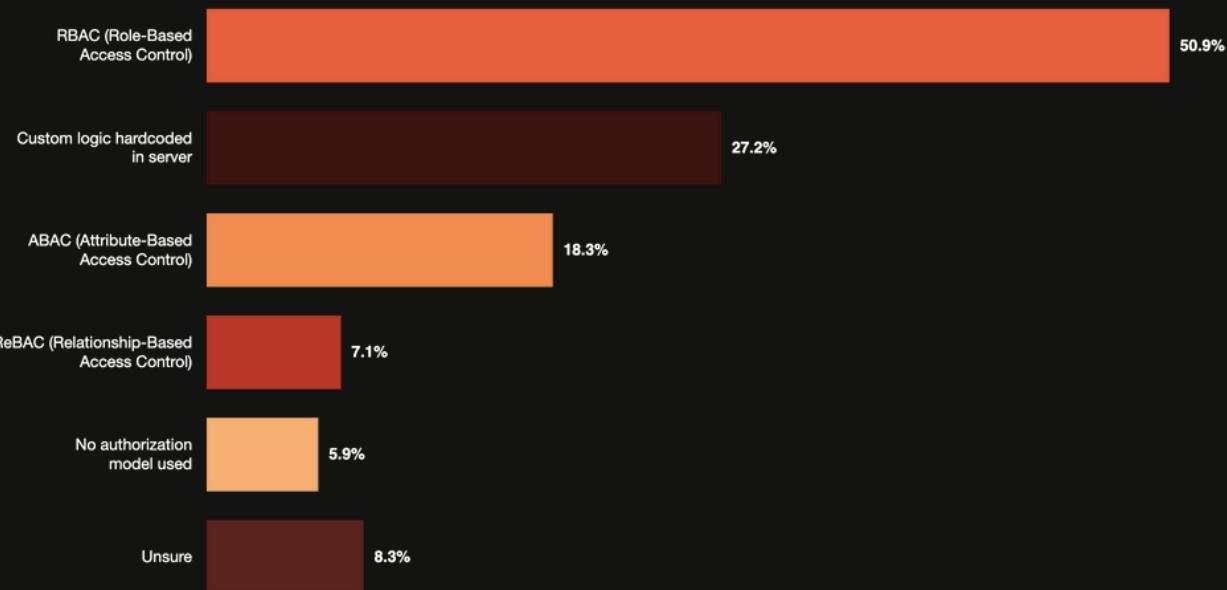
Authorization Is Often Hardcoded and Fragile

Custom authorization logic is widespread – and risky

While RBAC remains the industry standard, it is struggling to handle the dynamic, autonomous nature of agentic workflows. Organizations are increasingly relying on fragile, hardcoded logic or "shadow" authorization chains where agents create and task other agents without central oversight.

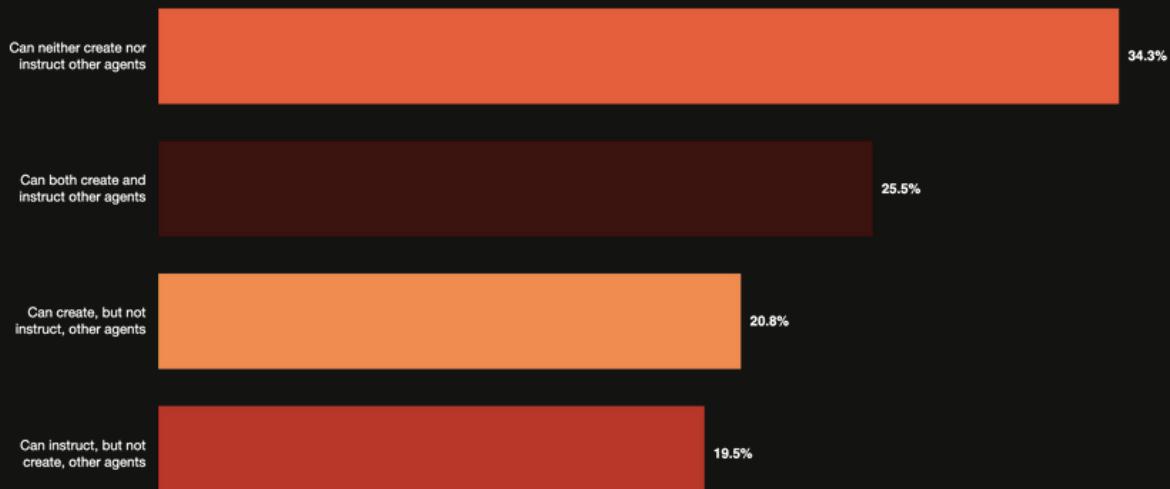
Half of technical teams rely on **RBAC**, but over a quarter (**27.2%**) have reverted to **custom, hardcoded logic** within servers to manage complex agent interactions, a method that is difficult to audit at scale.

Authorization Models for AI Agent & MCP Server Interactions



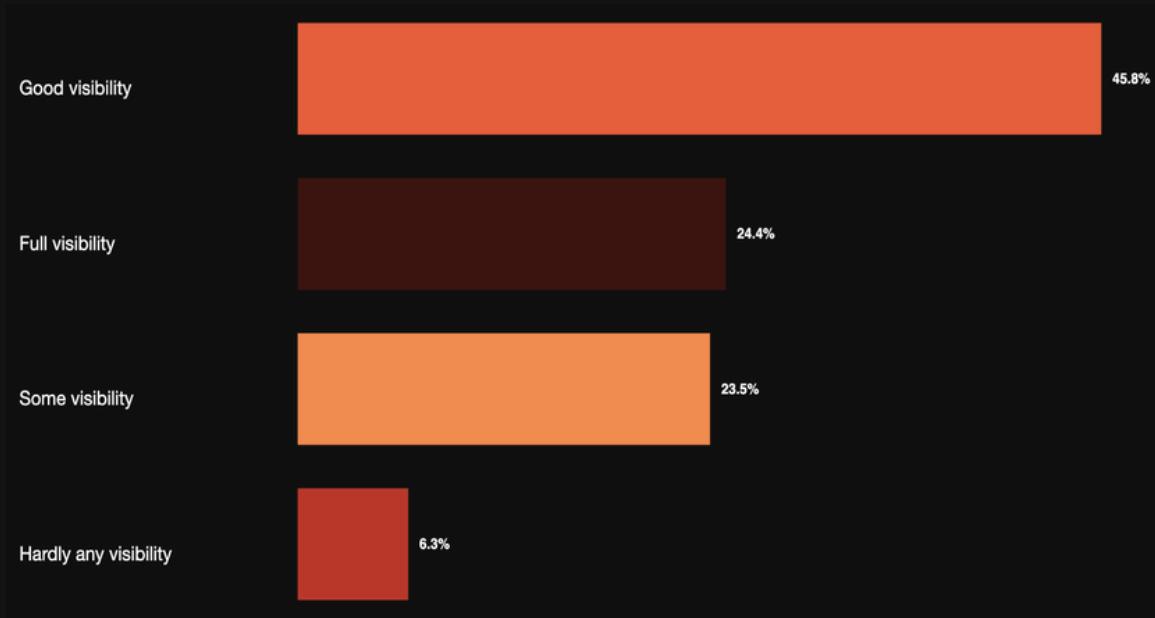
The autonomous Chains of Command: 25.5% of deployed agents are capable of both creating and instructing other agents, effectively establishing autonomous "chains of command" that may bypass traditional human-centric authorization gates.

AI Agent Autonomous Capabilities



The Visibility Gap: Only **24.4%** of organizations report having **full visibility** into which AI agents are interacting with others (A2A communication), leaving the majority of enterprises blind to how authority is being delegated internally.

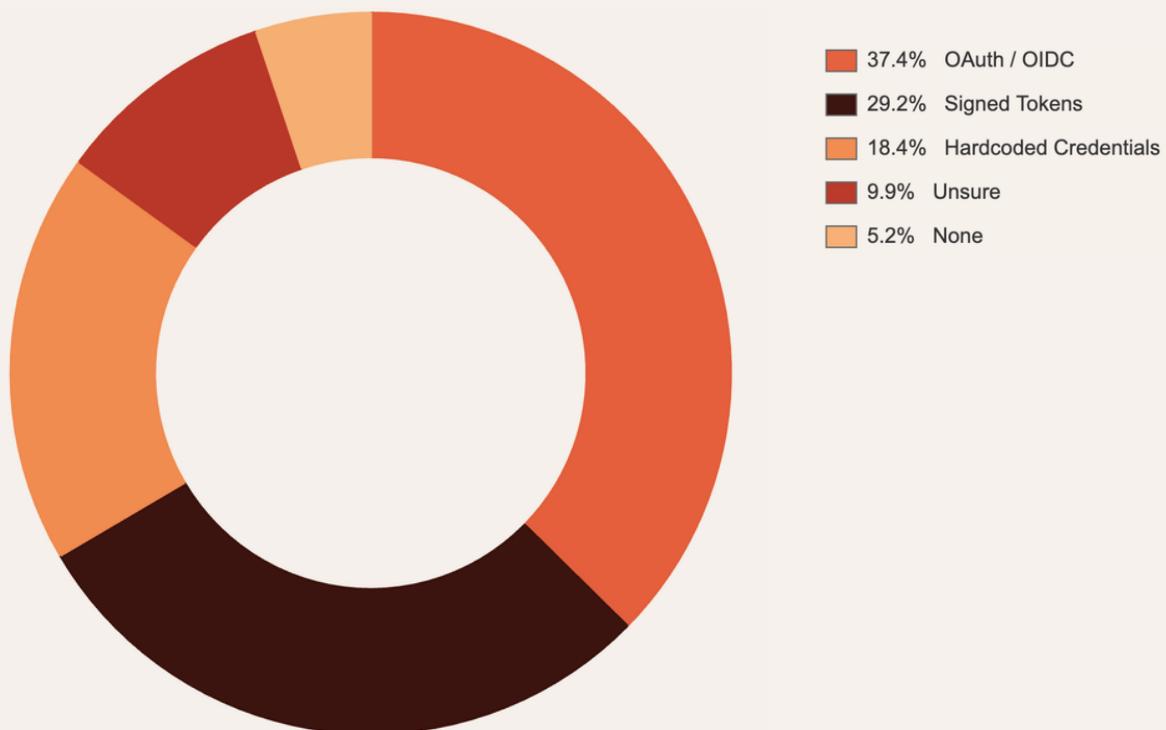
Visibility into AI agents interacting directly with other AI agents



Beyond the Data

As autonomous chains of command increase in complexity, **AI gateways** are emerging as a key tool for supporting AI governance, according to Gartner. Read the [Gartner® 2025 Market Guide for AI Gateways](#) to see why adoption is expected to reach 70% by 2028.

Tool and Upstream Authentication Methods for Agents and MCP Servers



Tool Access and MCP Are Where Risk Concentrates

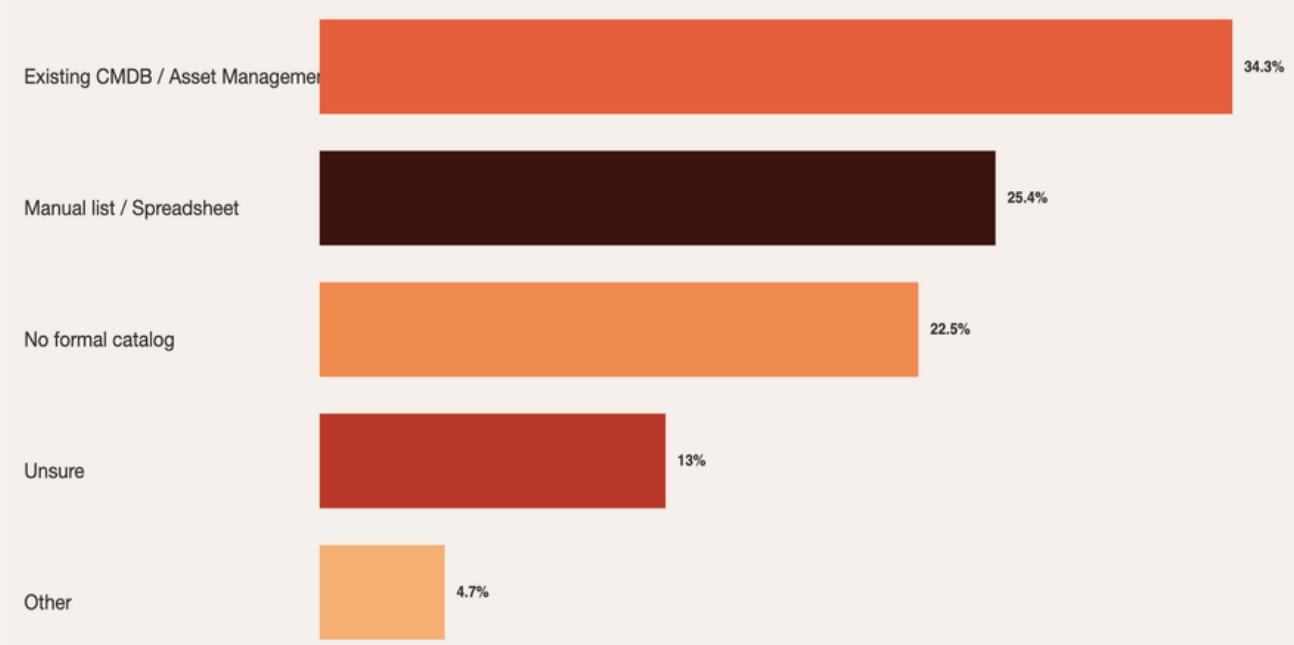
Tool authentication is inconsistent and often over-permissive

AI agents derive their power from their ability to interact with tools (databases, SaaS apps, internal APIs). However, this connectivity is creating a vast, unmapped attack surface. While the **Model Context Protocol (MCP)** is rapidly becoming the standard for this "plumbing," it remains dangerously disconnected from enterprise identity governance.

OAuth adoption is high (51.5%), but over a quarter of technical teams still rely on hardcoded credentials to connect agents to tools. Alarmingly, 7.1% of organizations use no authentication at all for these upstream connections.

AI agents are largely invisible to traditional asset management. **22.5%** of organizations have **no formal catalog** of their agents or MCP servers, and **25.4%** rely on manual spreadsheets that are outdated the moment they are saved.

How AI Agents and MCP Servers are Cataloged



Takeaway

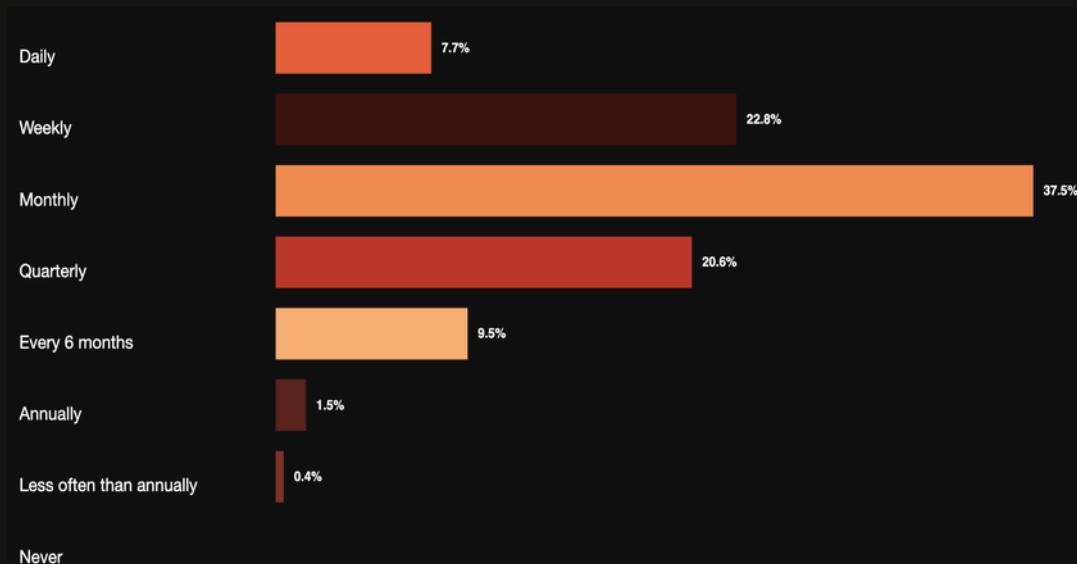
The tool layer is where agents touch real systems, and where weak authentication or delegation has immediate consequences. The rise of MCP provides a technical standard, but security teams must now layer identity-aware proxying over these connections to prevent them from becoming "shadow" backdoors.

Auditing Is Periodic, Not Continuous

💡 Most organisations review agent activity after the fact

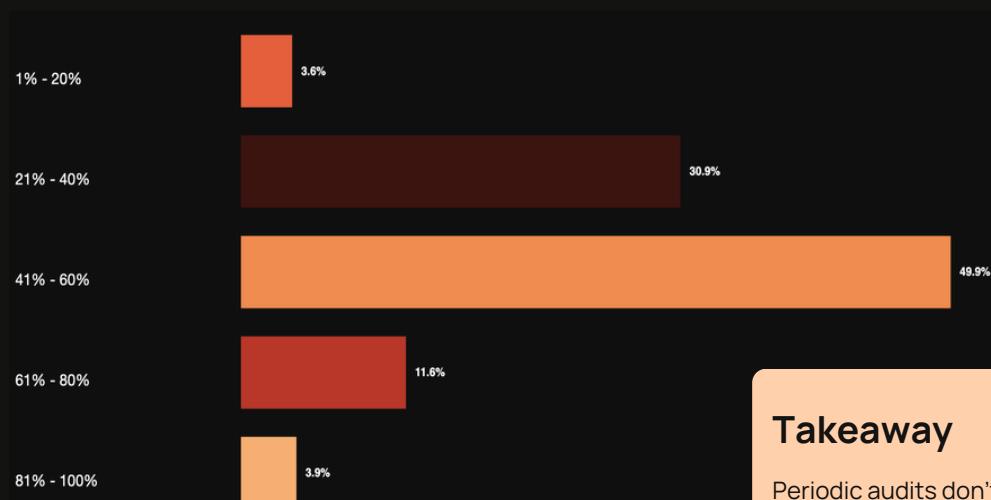
Most organizations review agent activity after the fact, creating a dangerous lag between an autonomous agent's actions and security detection. While AI agents can execute hundreds of tasks per second, only **7.7%** of organizations audit their activities daily. The majority (**37.5%**) rely on monthly reviews, leaving a significant window for undetected misuse or errors.

Frequency of AI Agent Audits by security or compliance teams



Visibility into live AI agents is very limited. Only **3.9%** of organizations report that more than 80% of their AI agents are actively monitored and secured. Nearly a third of organizations (**30.9%**) actively monitor and secure less than 40% of their deployed agent fleet. Technical builders are painfully aware of this deficit, with **57.4%** citing "insufficient observability (logging, monitoring, audit trails)" as a primary security concern when developing agentic systems.

Percentage of AI Agents Actively Monitored and Secured



Takeaway

Periodic audits don't match the speed or autonomy of agents. To eliminate real-time security blind spots, organizations must shift from compliance-based reviews to continuous, automated monitoring.

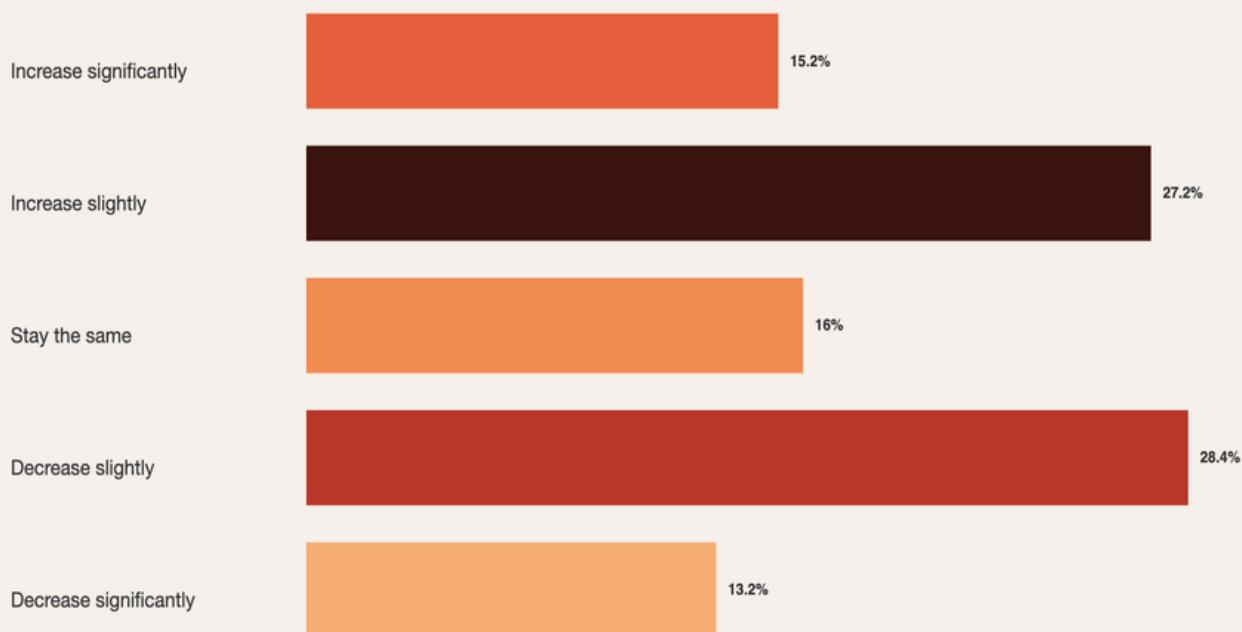
Investment and Regulation Create False Comfort

Risk is rising faster than security investment

Organizations are caught in a "wait-and-see" paradox. While executives express high confidence that current regulations (like the EU AI Act) mitigate agentic risk, technical budgets are not expanding to meet the unique security requirements of autonomous systems. This regulatory comfort is masking a significant funding gap.

Security spend is not keeping pace with adoption. According to the Executive Survey, nearly as many organizations expect their AI agent security investment to decrease (41.6%) as those who expect it to increase (42.4%) over the next 12 months.

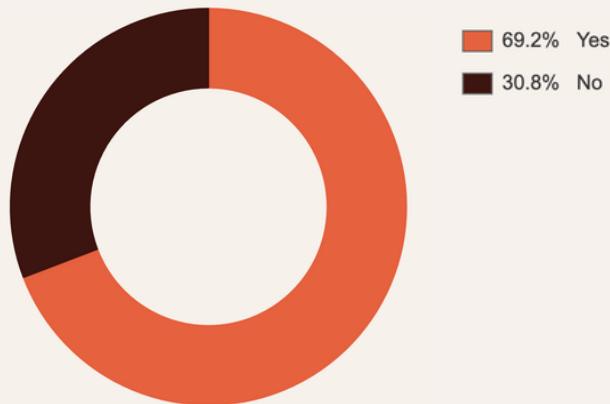
Expected Change in AI Agent Security Investment (Next 12 Months)



Despite the structural gaps in identity and authorization identified by technical teams, 69.2% of executives believe existing regulations are already sufficient to address the risks posed by autonomous agents.

Perception of Regulation Sufficiency - Do you believe current regulations (e.g. GDPR, EU AI Act) sufficiently address the risks posed by autonomous AI agents?

Technical teams remain more skeptical. Responses indicate that while "compliance" boxes are being checked, the actual implementation of agent security often relies on shared accounts and personal credentials to bypass budget-related friction.



Real-World Evidence

Practitioner stories

"Honestly, general LLM security is still a concern on an enterprise level so we have all been using our own personal accounts with the agents. Therefore, we haven't yet given much focus to agent security since we are still finalizing our building and workflows."

VP, Director, or Manager | Telecom, Media & Technology | 1000-10000 employees.

Practitioner stories

"We noticed that some agents were sharing passwords to access internal tools, which is a security risk. We addressed it by disabling shared accounts, creating individual logins, and conducting training on secure access practices."

VP / Director / Manager | Financial Services | 10,000+ employees

Practitioner stories

"During testing, we discovered that the agent was granted broader permissions than necessary due to a shared service account configuration. This meant that under certain prompt conditions, the agent could access endpoints outside its intended scope."

Principal / Lead | Financial Services | 1,000-10,000 employees

Practitioner stories

"During a project, we discovered that one of our machines had access keys to production servers that were not properly accounted for. This created a potential security risk."

Developer / Architect / Engineer | Financial Services | 10,000+

Practitioner stories

"Accessing infrastructure, using devops team high privilege access."

Developer / Architect / Engineer | Telecom, Media & Technology | 500-1,000 employees

Practitioner stories

"The root cause was a combination of factors: lack of fine-grained permission boundaries for agents, reuse of long-lived API keys, and insufficient visibility into agent decision paths when invoking tools."

Principal / Lead | Financial Services | 1,000-10,000 employees

Takeaway

Regulatory confidence is masking a funding crisis. Organizations are relying on existing laws to manage risk while failing to invest in the technical infrastructure, like dedicated agent identities and automated authorization, required to secure autonomous workflows.

What The Data Ultimately Tells Us

Across both surveys, one pattern dominates:

- AI agent security is an execution problem, not an awareness problem.

Organizations understand the risks, and as the data shows, incidents are already occurring in production environments.

What's missing is **cohesion**:

- consistent identity models
- centralized enforcement
- clear ownership
- continuous visibility

AI agents are already part of your infrastructure.
Security now has to catch up.