cloud security alliance®

token

Survey Report

Autonomous but Not Controlled

AI Agent Incidents Now Common in Enterprises

Acknowledgments

Lead Authors

Hillary Baron

Contributors

Marina Bregkou
Josh Buker
Ryan Gifford

Graphic Design

Stephen Lumpe

About the Sponsor

Token Security accelerates secure enterprise adoption of Agentic AI by discovering, managing, and governing every AI agent and non-human identity across the organization. From continuous visibility to understanding AI agent intent, enforcing least-privilege access, and managing AI agent identity lifecycles, Token Security provides complete control over AI and machine identities, eliminating blind spots, reducing risk, and ensuring compliance at scale.
https://www.token.security/

Table of Contents

Executive Summary

AI agents are rapidly becoming embedded across enterprise technology environments, operating across cloud platforms, internal systems, SaaS applications, and LLM-driven workflows. As these systems take on more autonomous roles, organizations are shifting from experimentation to governance—defining how agents are controlled, monitored, and integrated into existing security and risk frameworks. This report examines how organizations are structuring AI agent governance in practice, and where gaps remain as agent adoption scales.

1. Exception Handling Is Emerging as the Primary Control Plane for AI Agents

Organizations are not enforcing control continuously, but applying it at decision points. A majority operate agents autonomously for low-risk tasks with human review for higher-risk actions (53%), while only 13% report fully autonomous models. When agents exceed expected boundaries, actions are more likely to require approval (38%) or be logged (24%) than automatically blocked (11%). Monitoring is also largely periodic (59%), reinforcing a governance model based on checkpoints and escalation rather than real-time enforcement.

2. Organizations Believe They See Their AI Agents Yet Shadow Deployment Persists

While 68% report high confidence in their visibility into AI agents, 82% have discovered previously unknown agents in the past year. These agents most commonly appear in internal automation environments (51%) and LLM platforms (47%), which also align with where agents are actively deployed. This gap highlights a distinction between operational visibility and complete governance assurance, limiting the effectiveness of control models that depend on known and bounded agents.

3. AI Agent Lifecycle Controls Are Maturing But Decommissioning Lags

Organizations are improving front-end lifecycle practices, with 58% reporting clear documentation of agent purpose and 68% conducting permission reviews. However, only 21% have formal decommissioning processes, and just 19% express high confidence that agents are fully retired. This imbalance creates "retirement debt," where agents may persist beyond their intended use, retaining access and increasing long-term risk.

4. Risk and Delegation Are Becoming the Cornerstones of AI Agent Governance

Organizations are converging on action risk (63%) and human authorization (53%) as the primary signals for governing agent behavior. Nearly 79% view context-aware controls as important or very important, and 66% report clear guardrails defining agent boundaries. This reflects a shift away from static access models toward dynamic, context-driven policy decisions.

5. AI Agent Incidents Are Reshaping Operational Security Priorities

AI agent-related incidents are common, with 65% reporting at least one in the past year. These incidents have tangible business impact, including data exposure (61%) and operational disruption (43%), with no respondents reporting no impact. As a result, organizations are prioritizing monitoring (28%), risk management (29%), and permission control (19%), signaling a shift from discovery to managing agent behavior at scale.

Takeaway

AI agent governance is evolving into an interconnected system spanning visibility, lifecycle management, policy design, and monitoring. While organizations have established foundational controls, gaps remain in completeness, consistency, and end-of-life management. As agents expand in access and autonomy, governance must align with their operational impact. The next phase of maturity will depend on how effectively organizations integrate these capabilities into a cohesive model that can sustain control under real-world conditions.

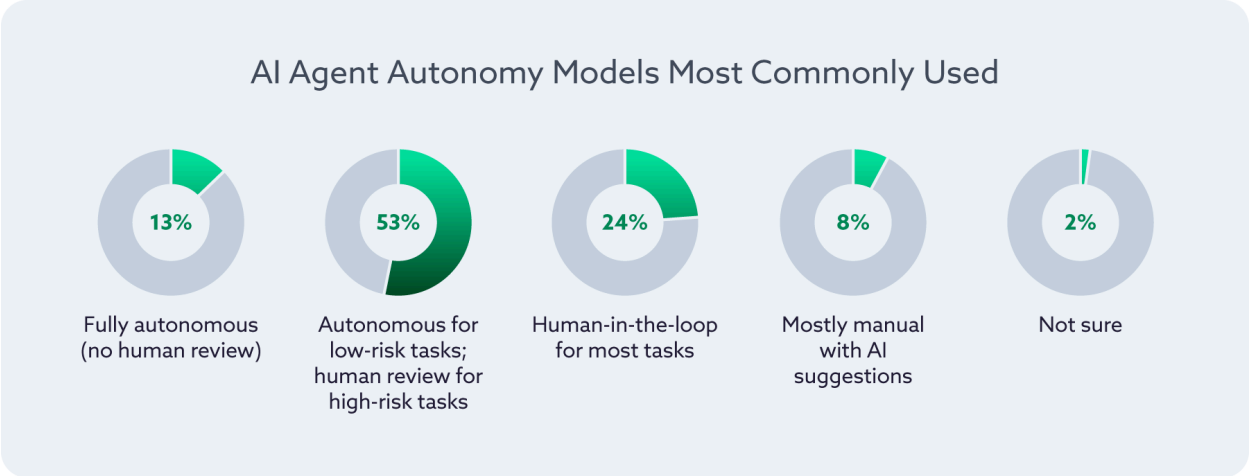
Key Findings

As AI agents become embedded in enterprise environments, organizations are developing new approaches to govern how these systems act, access resources, and interact with critical workflows. This includes defining how autonomy is controlled, how agents are identified and tracked, how their lifecycle is managed, and how risk is evaluated in real time. While progress is evident across these areas, the maturity of governance practices varies, revealing both emerging patterns and persistent gaps as organizations adapt to the operational realities of agentic systems.

Key Finding 1:

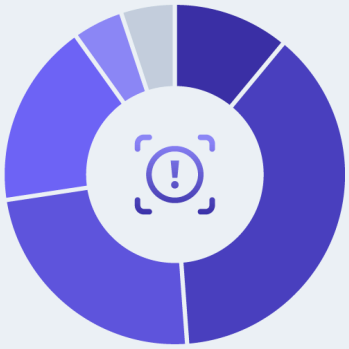
Exception Handling Is Emerging as the Primary Control Plane for AI Agents

Exception handling has emerged as the dominant operational control model for AI agents. A majority of organizations report that their AI agents operate autonomously for low-risk tasks but require human review for high-risk actions (53%), while an additional 24% rely on human-in-the-loop models for most tasks. Only 13% indicate fully autonomous deployments with no human review. This distribution indicates that autonomy is typically bounded and tiered, not absolute. Human control has not disappeared: instead it is concentrated on areas where the stakes are highest.



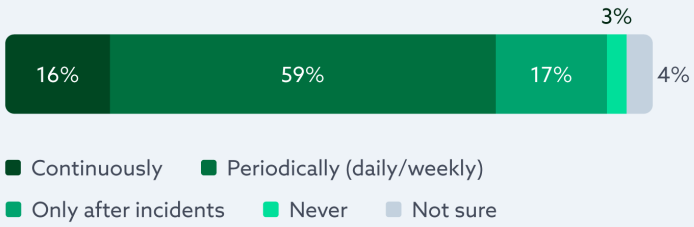
The way organizations respond when agents exceed their scope reinforces this structure. When an AI agent attempts an action outside its defined boundaries, the most common outcome is that the action requires human approval (38%). Another 24% allow the action to proceed, only log it for traceability, while 17% report that outcomes vary by system. Only 11% automatically block the action. The pattern reveals a clear preference for conditional intervention over automatic denial. When an agent attempts something unexpected, organizations are more likely to require approval or document the action than to stop it outright. Automatic blocking exists, but it is not the dominant approach.

What Typically Happens When AI Agents Attempt Actions Outside Their Expected Scope



- **11%** Action is automatically blocked
- **38%** Action requires human approval
- **24%** Action proceeds but is logged
- **17%** Outcomes vary by system
- **5%** Not sure
- **5%** N/A - do not have such controls

How Often AI Agents Are Monitored for Deviations from Expected Behavior



This conditional model extends beyond approval workflows into ongoing oversight. Most organizations monitor AI agents periodically, such as daily or weekly (59%), while just 16% report continuous monitoring. A notable 17% monitor only after incidents occur. The predominance of periodic oversight indicates that runtime governance is structured around checkpoints and review cycles instead of real-time, always-on enforcement. Oversight is present, but it is not universally embedded at every execution layer.

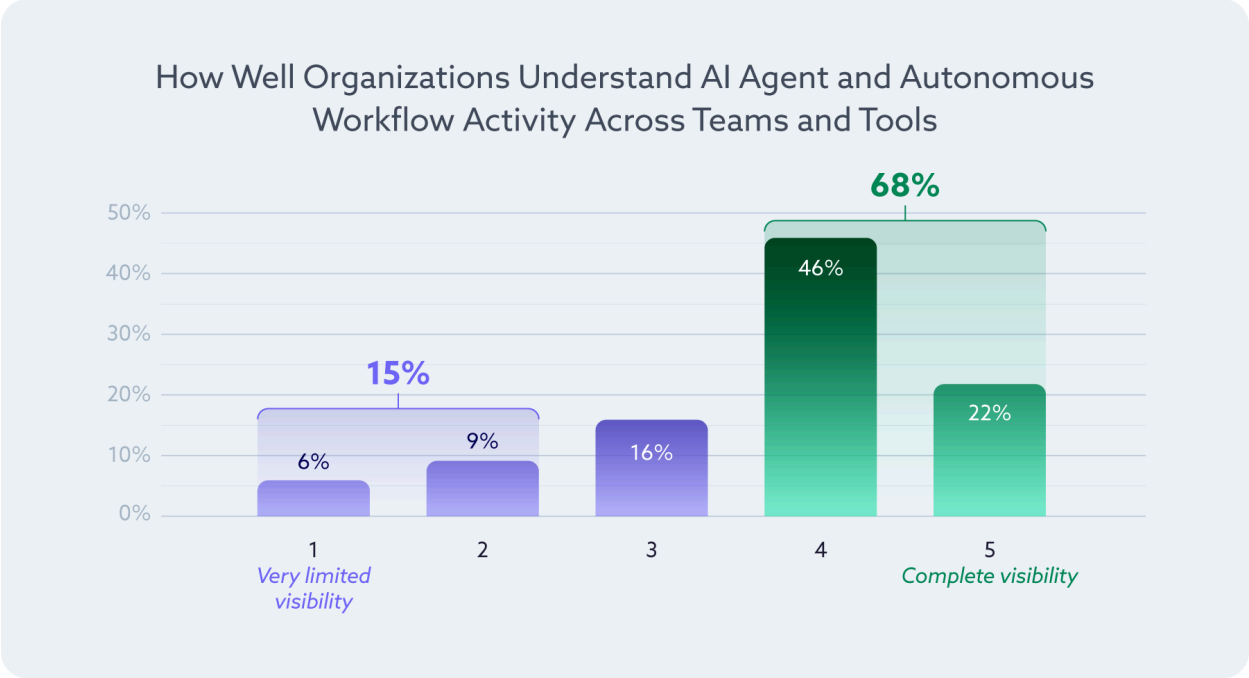
The data shows that most organizations are not trying to prevent every unexpected action in advance. Instead, they allow AI agents to operate within defined boundaries and step in when risk increases. Higher-risk or out-of-scope actions are routed to approval workflows, logged for review, or handled differently depending on the system. Control is applied at key moments rather than enforced uniformly at all times.

This approach depends on clearly defining what an agent is supposed to do. When an agent's purpose is well scoped, it can operate independently within that intent. When it moves beyond those boundaries, additional review or authorization is triggered. In this way, governance is built around defined limits and managed escalation, not constant restriction. As agents take on broader responsibilities, this model places greater weight on knowing when those limits are crossed. That raises a natural follow-on question: how well can organizations see all agents operating across their environments?

Key Finding 2:

Organizations Believe They See Their AI Agents Yet Shadow Deployment Continues

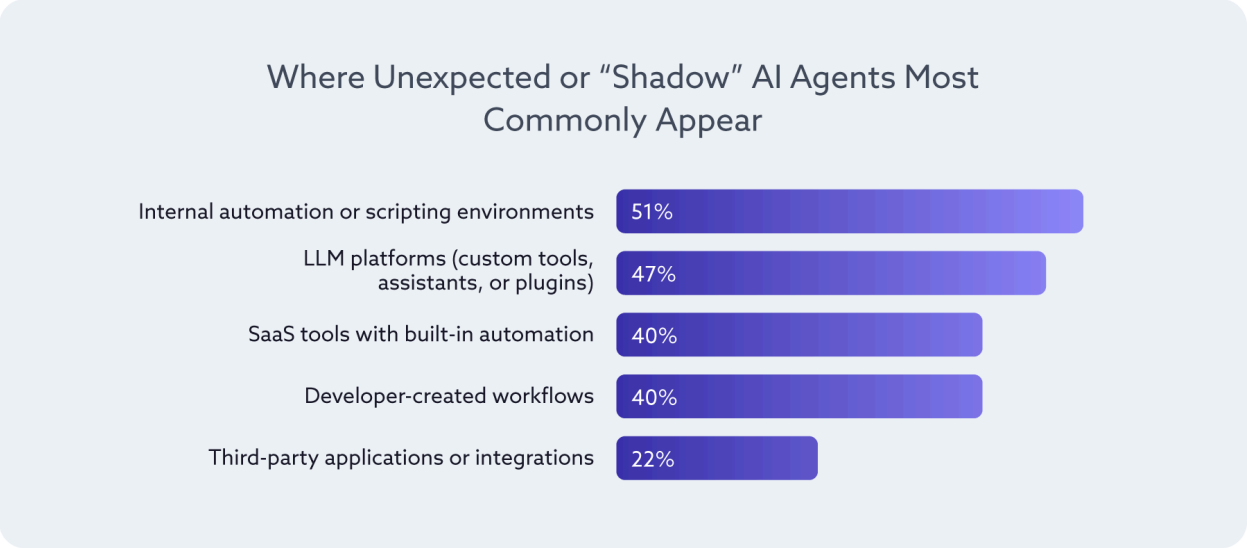
Organizations express strong confidence in their visibility into AI agents, yet frequent surprise discovery reveals a persistent gap between operational awareness and assurance-grade oversight. When asked to rate their understanding of all AI agents and autonomous workflows running across teams and tools on a five-point scale—where 5 represents complete visibility and 1 represents limited visibility—22% select 5, and 46% select 4. In total, 68% rate their visibility as high (4–5), while 15% report limited visibility (1–2). On the surface, this suggests that AI agents are sufficiently visible to support day-to-day operations and management. Most organizations believe they know what is running and where.



However, that confidence coexists with widespread surprise discovery of AI agents. In the past year, 82% report discovering at least one AI agent or autonomous workflow that was created without the knowledge of security, IT, or governance teams. Of those, 41% experienced this multiple times, while another 41% encountered it at least once. Only 11% report no such

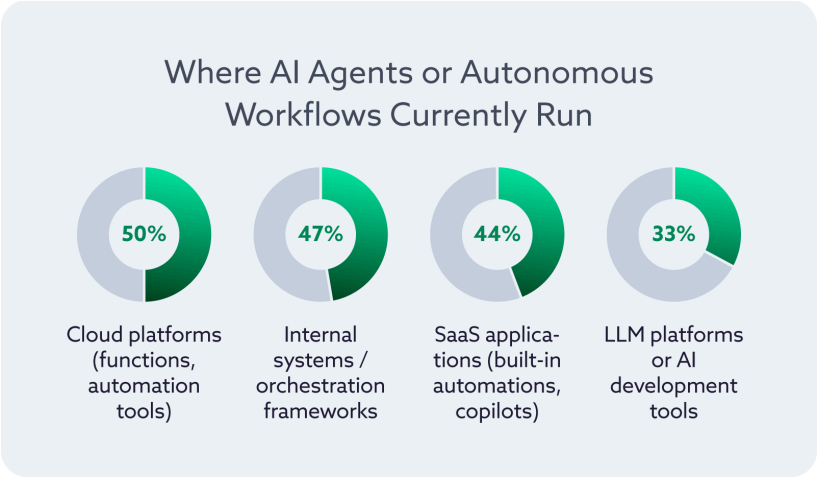
discoveries. This finding has direct implications for the exception-driven control model described earlier. Conditional intervention assumes that agents are known and operating within defined boundaries. When previously unknown agents surface, those boundaries may not exist or may not be consistently enforced. The result is a gap between perceived oversight and actual governance coverage.

Where these shadow agents appear helps explain why surprise discovery remains so common. Shadow agents most commonly emerge in internal automation or scripting environments (51%) and LLM platforms, including custom tools, assistants, and plugins (47%). SaaS tools with built-in automation (40%) and developer-created workflows (40%) also feature prominently.



These environments are built for speed and decentralized experimentation, allowing teams to create and deploy workflows quickly, often outside centralized oversight. This dynamic complicates the exception-driven control model described earlier: approval and escalation pathways are effective only when agents are known and within defined governance boundaries. As deployment expands across cloud platforms, internal systems, SaaS applications, LLM tools, and MCP-connected environments, operational visibility may be sufficient for daily management, but achieving complete, assurance-grade oversight becomes more difficult.

This concentration of shadow agents is not occurring at the margins of the environment. It closely mirrors where AI agents are most actively deployed. Half of organizations report agents running in cloud platforms (50%), 47% in internal systems or orchestration frameworks, 44% in SaaS applications, and 33% in LLM platforms or AI development tools. In other words, the environments



enabling legitimate agent adoption are the same ones producing unexpected discovery or shadow agents. As AI agents expand across internal automation frameworks, SaaS ecosystems, LLM platforms, and MCP-connected architectures, visibility challenges do not diminish—they move deeper into core operational systems.

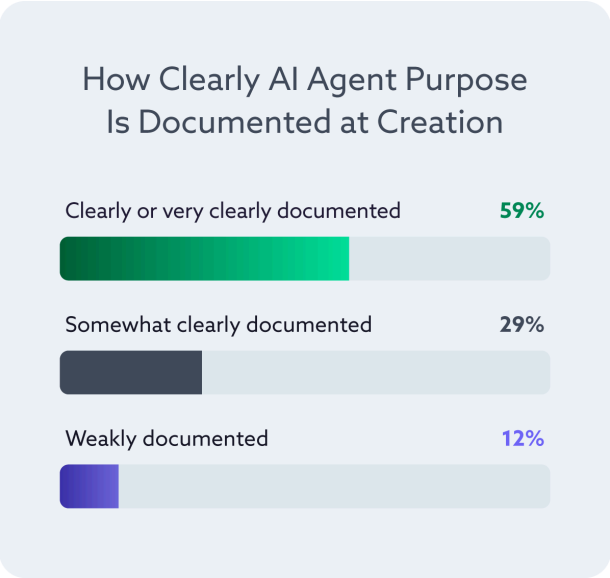
The coexistence of high self-rated visibility and frequent surprise discovery highlights a structural gap between operational awareness and governance assurance. Operational visibility allows teams to manage known agents effectively. Governance assurance, however, requires confidence that all agents—authorized or not—are identified, scoped, and subject to control pathways. The data suggests many organizations have sufficient visibility to operate AI agents day to day, but not enough completeness to guarantee that exception-driven controls apply universally.

As agent ecosystems expand across cloud platforms, internal orchestration systems, SaaS applications, LLM environments, and MCP-connected architectures, this distinction becomes more consequential. Exception-based governance depends on agents being known and bounded. When visibility is incomplete, those safeguards may not consistently engage. Closing that gap requires stronger lifecycle discipline to ensure agents remain identified, authorized, and properly retired over time—a challenge that becomes central in the next finding.

Key Finding 3:

AI Agent Lifecycle Controls Are Maturing But Decommissioning Lags

Organizations are strengthening early-stage AI agent lifecycle controls, but weak decommissioning practices are creating a growing layer of "retirement debt." A key part of this front-end discipline is clarity of intent at creation. Fifty-nine percent report that an agent's intended purpose is clearly or very clearly documented, while only 12% indicate weak documentation. Defining what an agent is meant to do establishes the governance boundary within which it can safely operate.

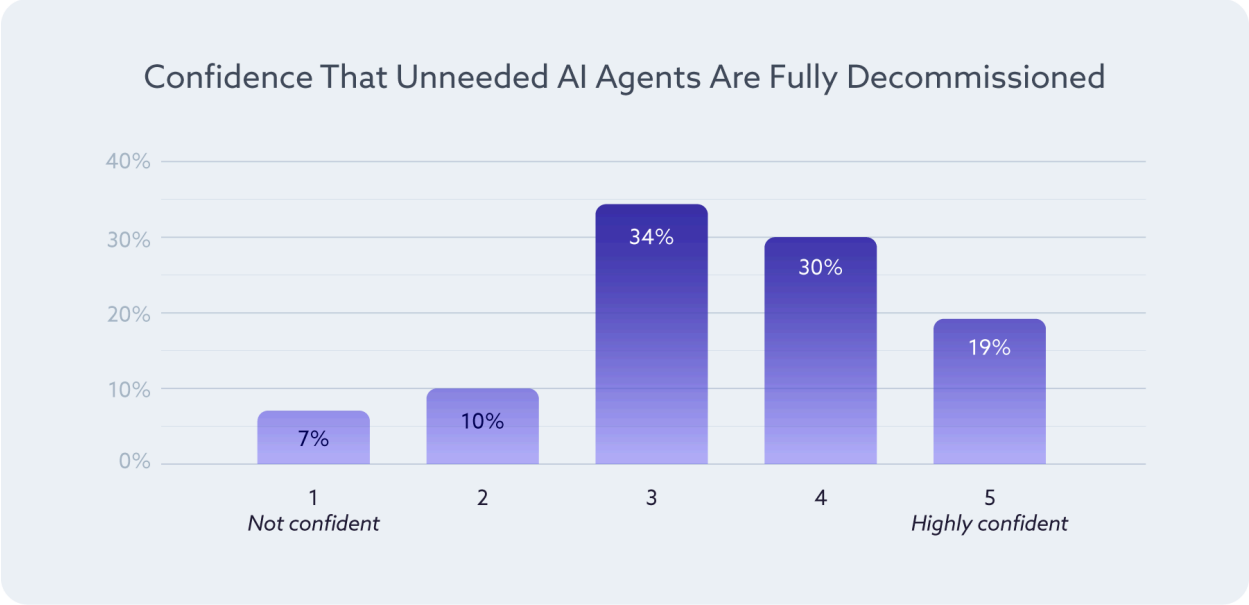


That clarity is reinforced by more formal lifecycle controls. A majority report conducting periodic permission reviews (68%), and over half have defined creation or onboarding processes (52%). Nearly half monitor agent behavior or performance (44%), and 35% document ownership. Defined onboarding, permission reviews, and documented ownership reflect an effort to bring agents within known governance boundaries. As adoption increases, the focus is shifting from experimentation to structured oversight, at least at the point of creation.

However, end-of-life governance is notably less developed. Formal decommissioning processes are reported by only 21%, making it the least adopted lifecycle step by a significant margin. While onboarding and permission reviews are common, structured retirement remains comparatively rare. The imbalance

suggests that organizations are more focused on enabling and maintaining agent functionality than on systematically determining when an agent should be removed.

The limited adoption of formal decommissioning processes is echoed in respondents' confidence levels. When asked to rate their confidence on a five-point scale—where 5 indicates high confidence that agents are fully removed, access is revoked, and keys are retired—only 19% select the highest rating. Confidence clusters at moderate levels instead, with 34% selecting 3 and 30% selecting 4. Seventeen percent report low confidence (1–2). This distribution suggests that even where decommissioning occurs, assurance that it is executed completely and consistently is not universal.



Together, these patterns point to the emergence of retirement debt: AI agents that may persist beyond their intended purpose, retaining credentials, permissions, or operational hooks after their business value has diminished. Unlike provisioning risk, which is visible at creation, retirement risk accumulates quietly over time. Agents that are well-documented and periodically reviewed can still outlive their mandate if no formal trigger or enforced process governs their removal.

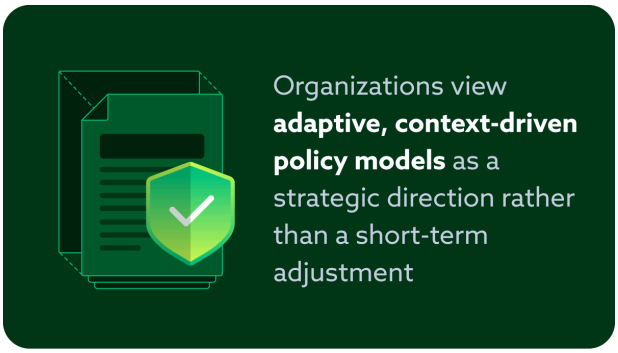
As AI agent populations scale across cloud platforms, internal orchestration systems, LLM environments, and MCP-connected architectures, unmanaged retirement becomes more than an operational oversight—it becomes a structural exposure. This lifecycle imbalance sets the stage for a broader governance challenge: how organizations define risk signals and delegation boundaries for agents that continue to operate with expanding scope and autonomy.

Key Finding 4:

Risk and Delegation Are Becoming the Cornerstones of AI Agent Governance

Risk and delegation are emerging as the primary policy signals shaping how organizations govern AI agent behavior. Underlying this shift is a broader pattern: AI agent risk is driven by the combination of access and autonomy—what systems an agent can reach and how independently it can act. As both increase, so does potential impact, making risk a central input into governance decisions. Sixty-three percent identify action risk level as a governing signal, and 53% cite user delegation or approval. By contrast, fewer point to data sensitivity (37%), environmental context such as time or system (34%), or agent behavior history (21%). This distribution indicates convergence around a focused set of decision criteria. Rather than fragmenting across many possible inputs, organizations appear to be standardizing on risk magnitude and human authorization as the most authoritative signals for policy evaluation.

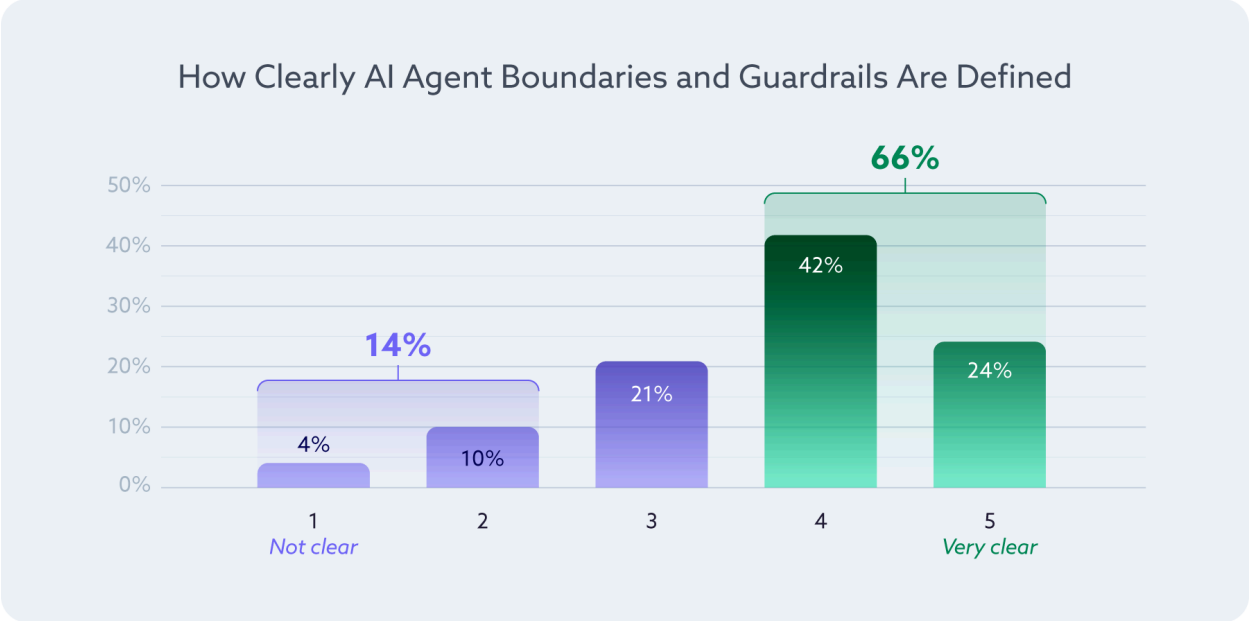
This emphasis reflects a shift away from static, role-based access models toward more adaptive decision frameworks. Risk level introduces dynamic evaluation—what is the potential impact of this action? Delegation introduces accountability—has a human explicitly authorized or scoped this action? Together, these signals create a policy language grounded in both consequence and intent. Static permissions may still define baseline access, but contextual evaluation increasingly determines whether high-impact actions proceed.



This emphasis on risk and human authorization is not limited to current practice; it is shaping future priorities. When asked to rate the importance of moving toward more context- or intent-aware controls over the next two years on a five-point

scale—where 5 indicates "very important"—47% select 4 and 32% select 5. Only 8% rate this shift as unimportant (1–2). The data suggests that organizations view adaptive, context-driven policy models as a strategic direction rather than a short-term adjustment.

For risk- and delegation-based policy to function effectively, agents must operate within clearly defined boundaries. Sixty-six percent report that the guardrails defining what an AI agent is allowed to do are clear or very clear (4–5), while 14% indicate limited clarity (1–2). This suggests that many organizations are investing in defining acceptable behavior at the outset—building on earlier improvements in purpose documentation and lifecycle controls. Clear AI agent intent establishes the scope within which agents can act; contextual signals such as risk level and human authorization determine when actions remain within that scope and when additional approval is required.

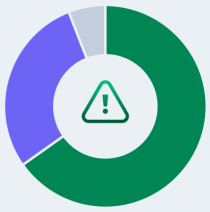


Taken together, the data indicates the emergence of a shared governance model for AI agents—one that centers on evaluating risk and validating delegation rather than relying exclusively on static permissions. As agent autonomy expands across cloud platforms, internal orchestration systems, LLM environments, and MCP-connected ecosystems, this convergence around contextual policy signals becomes foundational. The remaining question is how these evolving policy frameworks respond when incidents occur and theoretical controls are tested under real operational pressure.

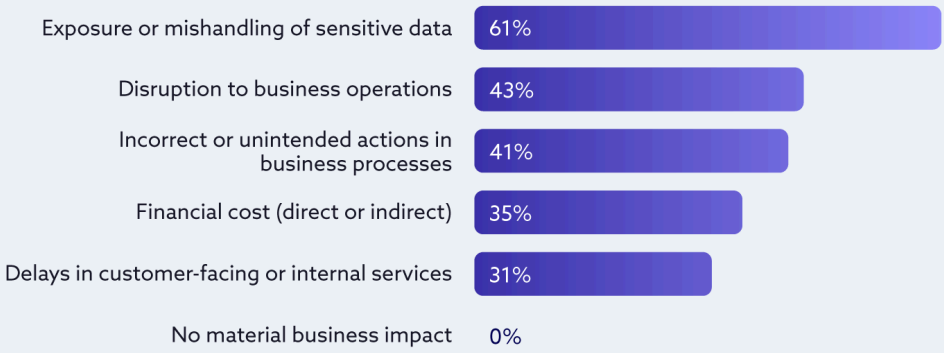
Key Finding 5:

AI Agent Incidents Are Reshaping Operational Security Priorities

Security incidents involving AI agents are widespread across organizations. Nearly two-thirds of organizations report experiencing a security incident involving an AI agent or autonomous workflow within the past 12 months (65%). Only 29% report no such incident, and the remainder were unsure or did not respond. This indicates that agent-related security events are now common operational realities rather than edge cases. AI agents have moved firmly into production environments, and with that shift comes measurable exposure.



These incidents affect multiple areas of the business and carry meaningful consequences. Among organizations that experienced an incident, 61% report exposure or mishandling of sensitive data, 43% report disruption to business operations, and 41% cite incorrect or unintended actions within business processes. Financial cost is reported by 35%, and 31% experienced delays in customer-facing or internal services. Notably, 0% report no material business impact. These outcomes show that AI agent incidents are affecting core enterprise functions, including data protection, operational continuity, financial performance, and service delivery. For organizations, this shifts AI agent governance from a technical oversight issue to a business risk management concern. Agent behavior must now be integrated into broader security, compliance, and operational resilience strategies rather than managed as an isolated automation challenge.



This incident exposure is shaping how organizations define their most pressing challenges. Managing risk or preventing unintended actions is identified as the top challenge by 29%, closely followed by monitoring agent activity and behavior at 28%. Controlling or reviewing agent permissions accounts for 19%, while only 13% cite understanding where agents exist. The distribution of challenges suggests that organizations are moving beyond basic inventory concerns and grappling with how to manage agent behavior at scale. Monitoring activity and controlling permissions now rival or exceed concerns about simply identifying where agents exist. The central difficulty is no longer awareness alone, but maintaining control over dynamic, autonomous systems operating within defined boundaries.

This pattern raises a structural tension. Organizations report frequent, material incidents while relying largely on periodic oversight models rather than continuous monitoring (see q18, prior key finding). As agent activity scales and actions occur at machine speed, checkpoint-based supervision may leave gaps between detection and impact. The combination of high incident prevalence and delayed visibility increases pressure to strengthen real-time monitoring, tighten permission controls, and reduce the window between deviation and response.

Taken together, the findings suggest that incident experience is accelerating a broader shift in how organizations govern AI agents. Exception-driven control models, improving visibility, and stronger front-end lifecycle practices provide a foundation, but confirmed incidents expose where those safeguards are uneven or incomplete. When agents operate beyond defined boundaries, remain active past their intended purpose, or act without timely oversight, the business impact becomes immediate.

For organizations, this implies that AI agent governance can no longer be treated as a collection of isolated controls. Visibility, lifecycle discipline, contextual policy signals, intent-based permissioning, ongoing privilege optimization, and monitoring cadence must function as an integrated system. As agents scale across cloud platforms, internal systems, SaaS applications, LLM environments, and MCP-connected architectures, governance models must align with the speed and autonomy of the systems they oversee. The maturation underway is not simply about adding more controls—it is about ensuring that boundaries, delegation, monitoring, and retirement processes operate cohesively under real operational pressure.

Conclusion

AI agents are now embedded in core enterprise systems, and organizations are actively building governance models to manage their autonomy. The data shows meaningful progress: autonomy is typically bounded rather than absolute, policy decisions increasingly factor in risk and explicit human authorization, and early-stage lifecycle controls are becoming more structured.

Yet these elements do not operate in isolation. Exception-driven control depends on complete visibility. Visibility depends on disciplined onboarding and ownership. Context-aware policy depends on clearly defined intent of AI agents. Monitoring must align with the speed and scale of agent activity. When any one layer weakens—whether through shadow deployment, uneven retirement practices, poor access control, or delayed detection—the effectiveness of the overall model is reduced.

AI agent risk is best understood through the combination of access and autonomy. Agents with limited access and strong human oversight typically present lower risk. As access expands to business-critical systems and autonomy increases, the potential impact grows significantly. This creates a clear prioritization model: agents with both high access and high autonomy require the strongest governance and monitoring. As organizations scale AI agents, this relationship becomes central to how risk is assessed and managed.

Taken together, these patterns point to a broader shift: AI agent governance is not a single control, but a system that must function cohesively across the agent lifecycle. Organizations appear to be moving in this direction, but maturity remains uneven. Front-end discipline is stronger than end-of-life management. Operational awareness is stronger than assurance-grade completeness. Policy intent is advancing faster than enforcement consistency.

In practice, this system depends on several foundational capabilities:

- **Maintain visibility across AI agents** – ensure agents operating across SaaS platforms, internal systems, and LLM environments are identified and within governance scope
- **Define and document agent purpose** – establish intended function to set operational boundaries and align access with that scope
- **Apply lifecycle governance consistently** – extend onboarding, ownership, review, and decommissioning processes across the full agent lifecycle
- **Evaluate actions based on risk and authorization** – use contextual signals such as action risk and explicit human approval to guide decision-making
- **Align monitoring with agent activity** – evolve from periodic oversight toward more continuous or event-driven detection models
- **Incorporate agents into enterprise risk models** – treat AI agents as part of broader security, compliance, and operational resilience frameworks

Confirmed incidents underscore the stakes. When agents deviate, persist beyond purpose, or operate outside defined boundaries, the impact reaches core business functions. As a result, AI agent governance can no longer be treated as a discrete technical domain. It is becoming part of enterprise risk management and operational resilience.

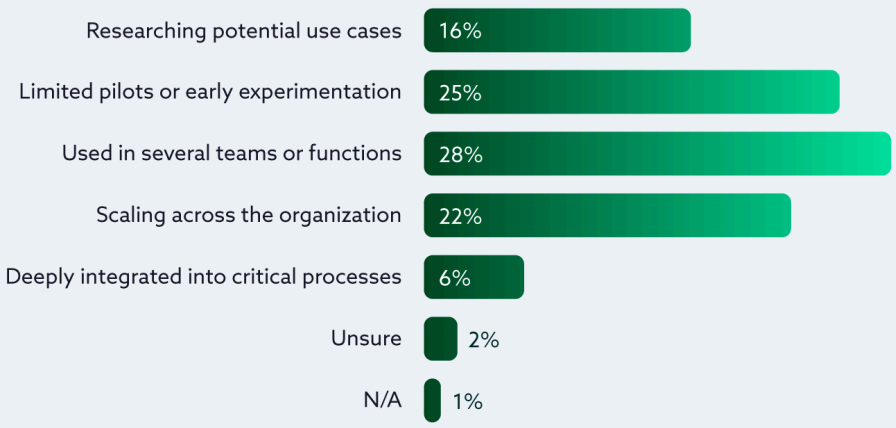
For organizations, the path forward is not simply adding more controls. It is ensuring that these capabilities operate together as a coherent system—one aligned to the speed, scale, and autonomy of modern agentic systems.

AI agents are scaling. Governance must scale with them.

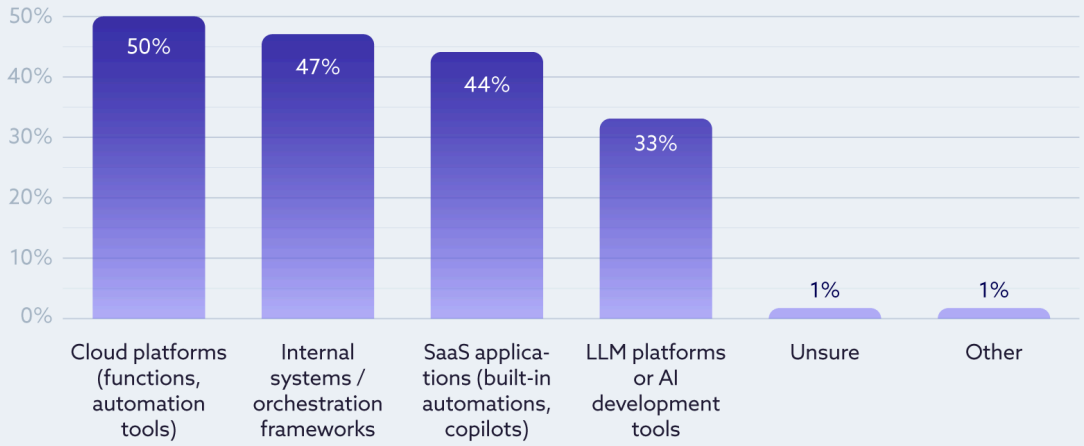
Full Results

Use and Adoption

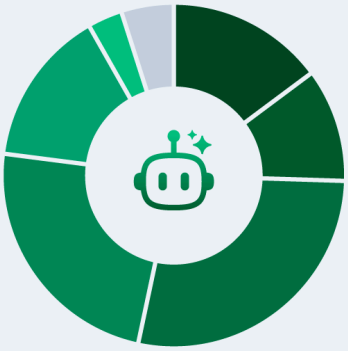
Which of the following best describes your organization's current use of autonomous or agentic AI?



Where do your AI agents or autonomous workflows currently run?



Approximately how many AI agents, automated copilots, or autonomous workflows do you believe your organization currently has in operation?

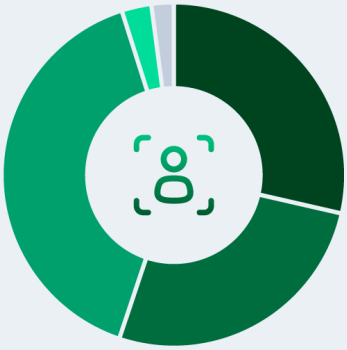


15%	Fewer than 10	**15%**	500–1,000
11%	11–50	**3%**	1,001+
28%	51–200	**5%**	Unsure
24%	201–500		

On average, how many systems or platforms do your AI agents connect to?

Agentic Identity Lifecycle Maturity

Does your organization define a clear "owner" or responsible person when a new AI agent or automated workflow is created?



29% Always

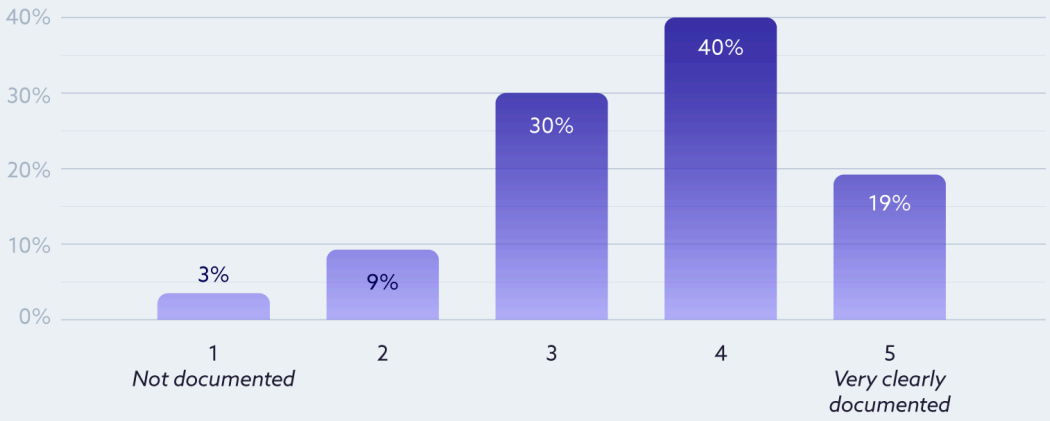
26% Usually

40% Sometimes

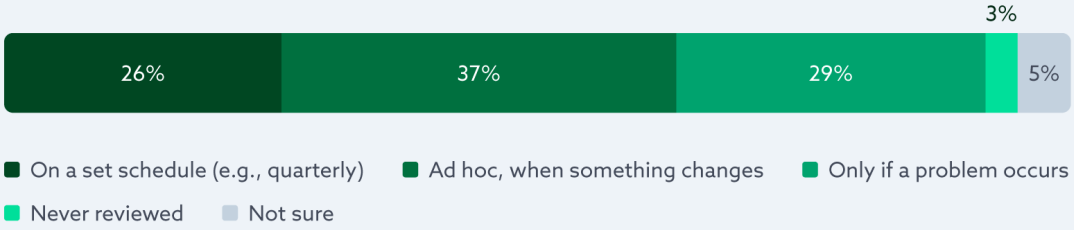
3% Never

2% Unsure

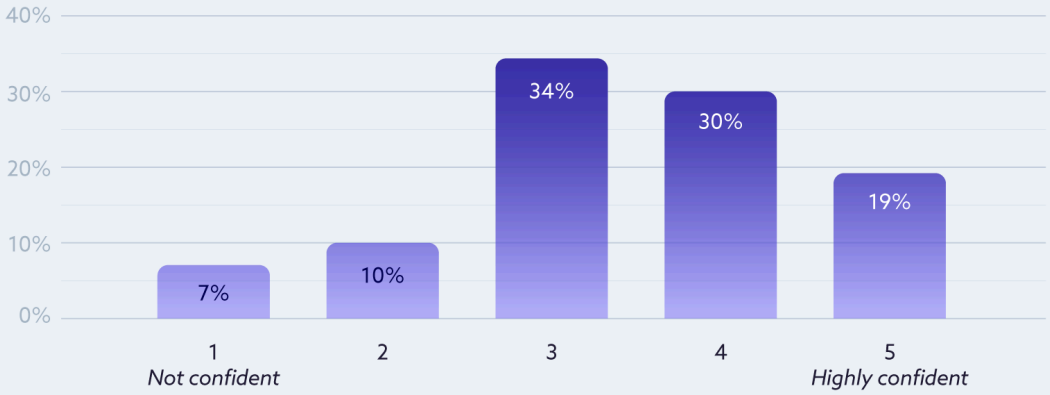
When AI agents are created, how clearly is their intended purpose documented?



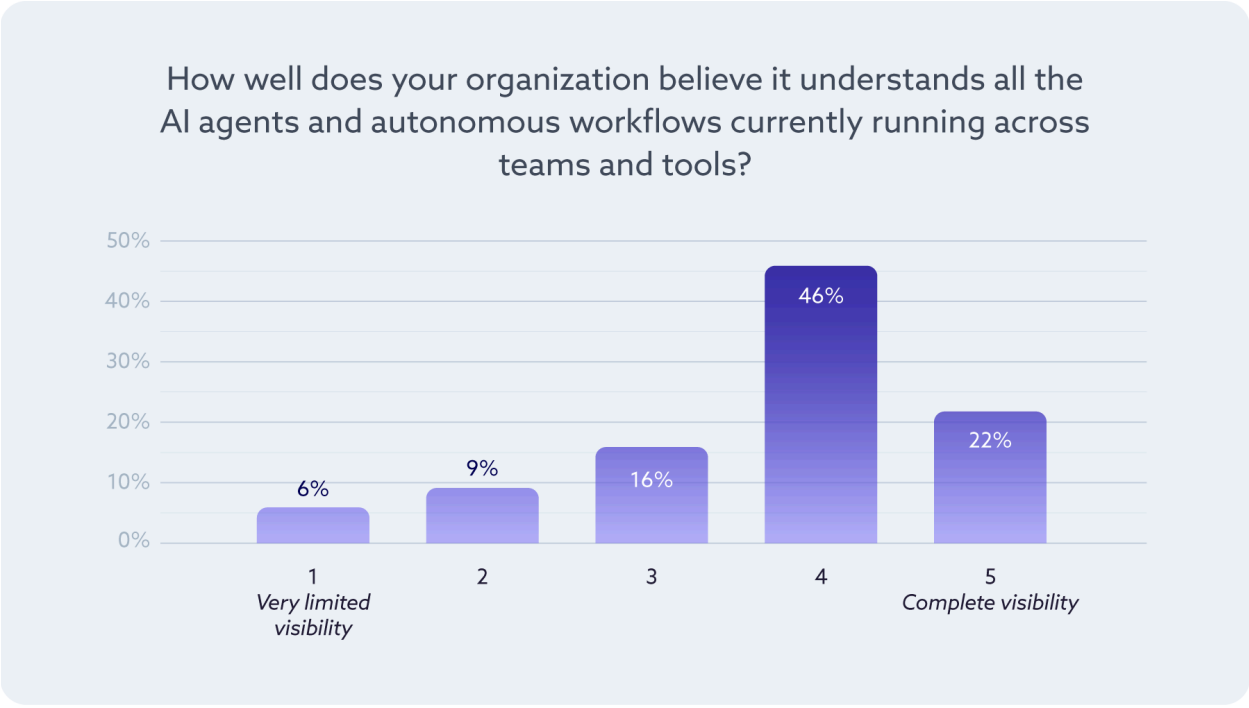
How often are an agent's permissions reviewed after initial deployment?



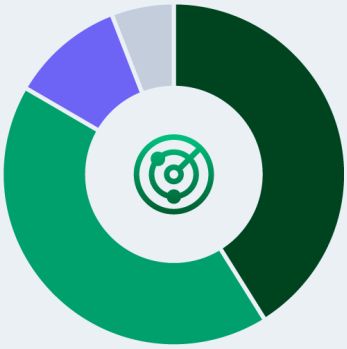
How confident are you that agents that are no longer needed are fully decommissioned (e.g., removed, access revoked, keys retired)?



Shadow Agent and Workflow Visibility



In the past year, has your organization discovered AI agents or autonomous workflows that were created without the knowledge of security, IT, or governance teams?



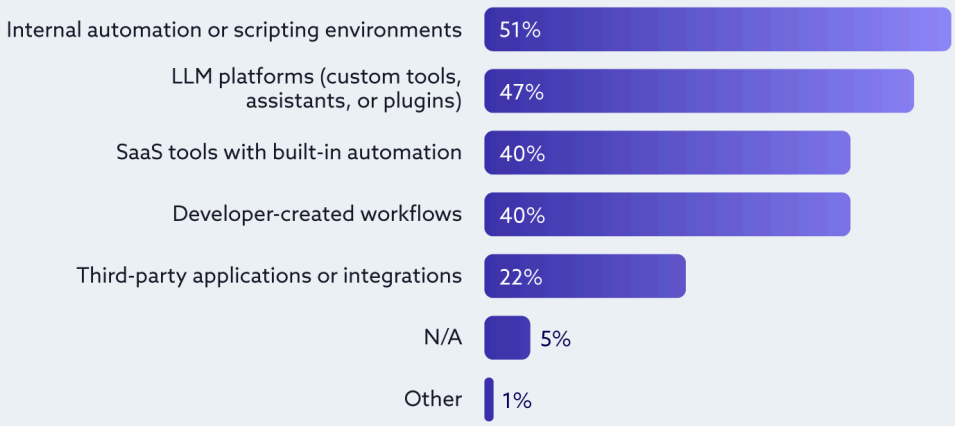
41% Yes, multiple times

41% Yes, at least once

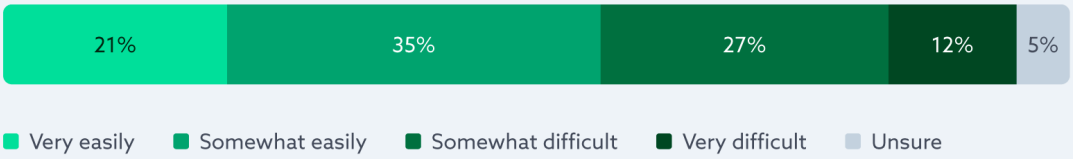
11% No

6% Unsure

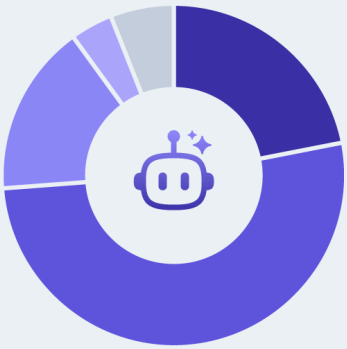
Where do unexpected or "shadow" agents most commonly appear?



How easily can your organization trace an action performed by an AI agent back to its triggering source (human or system)?



How frequently do AI agents interact with or trigger other agents or automated systems in your environment?

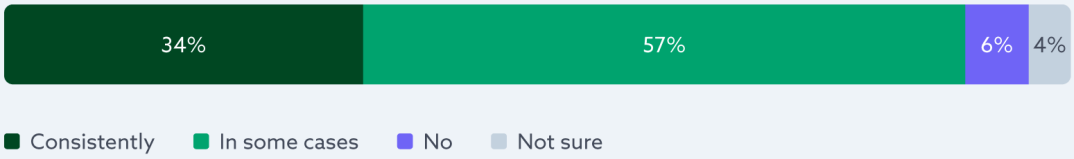


22% Frequently
52% Occasionally
16% Rarely
4% Never
6% Not sure

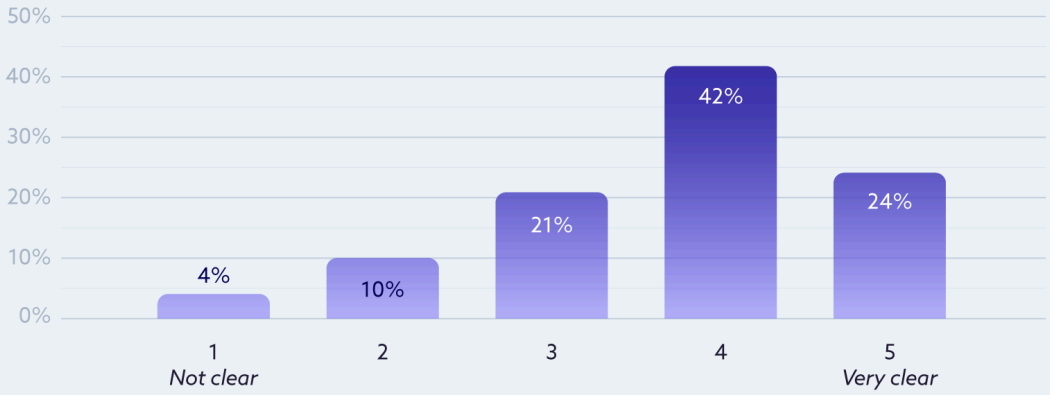
Intention, Autonomy, and Adaptive Guardrails

Which autonomy model best describes most of your AI agents?

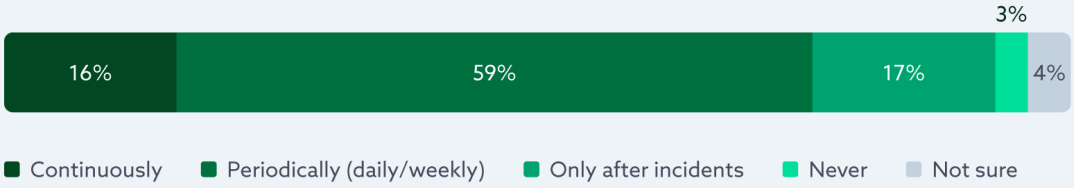
Do you scope AI agent privileges based on the agent's intended purpose or task?



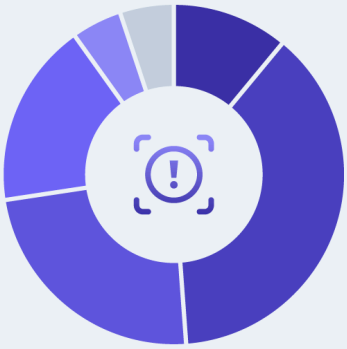
How clear are the boundaries or guardrails that define what an AI agent is allowed to do?



How often does your organization monitor AI agents for behavior that deviates from expected patterns?



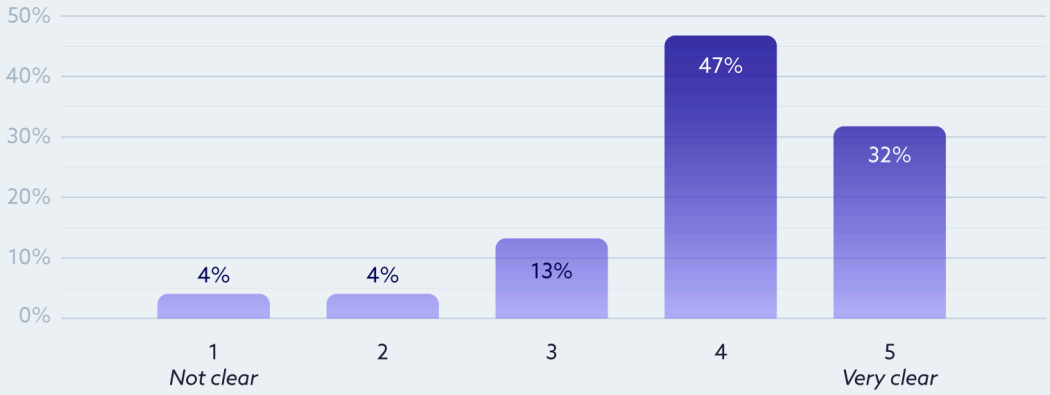
When an AI agent attempts an action outside its expected scope, what typically happens?



- **11%** Action is automatically blocked
- **38%** Action requires human approval
- **24%** Action proceeds but is logged
- **17%** Outcomes vary by system
- **5%** Not sure
- **5%** N/A - do not have such controls

Which context signals influence whether an AI agent is allowed to perform a sensitive action?

How important is it for your organization to move toward more context-aware or intent-aware controls for AI agents over the next two years?



Which of the following best describes your organization's biggest challenge with autonomous or agentic AI today?

29% Managing risk or preventing unintended actions

28% Monitoring agent activity and behavior

19% Controlling or reviewing agent permissions

13% Understanding where agents exist

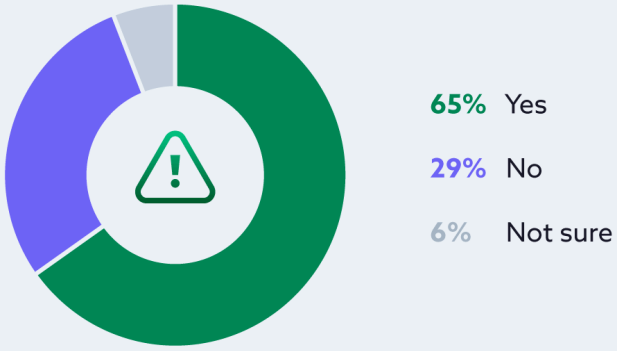
6% Assigning ownership and accountability

4% Lack of tools or expertise

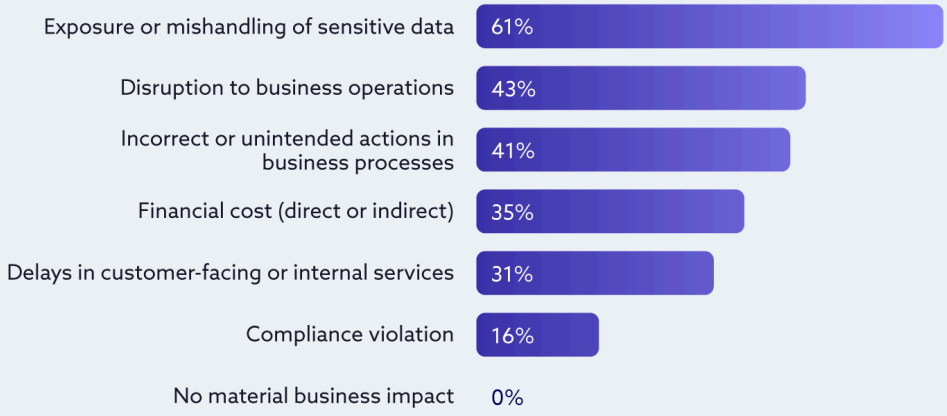
1% N/A - no challenges currently

Security Incidents

Has your organization experienced a security incident involving an AI agent or autonomous workflow in the past 12 months?

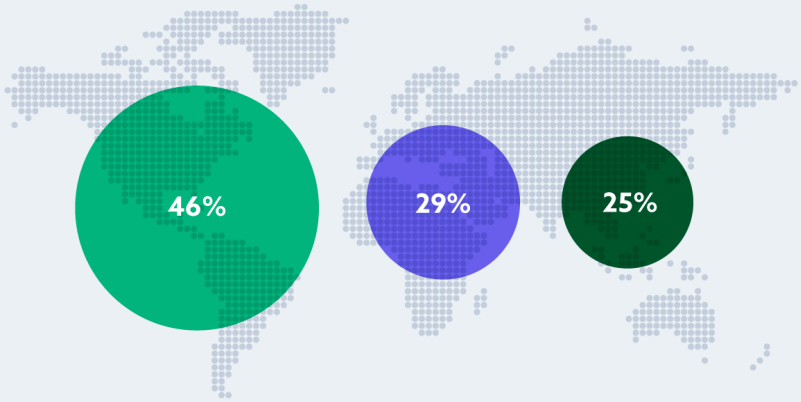


What impact did this security incident have on your organization?

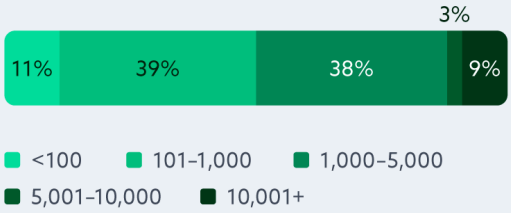


Demographics

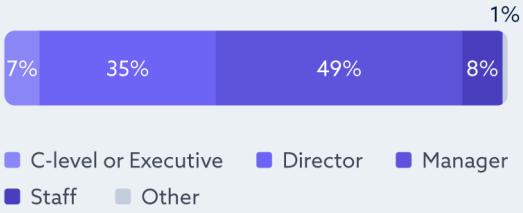
Location



Organization Size



Job Level



Primary Functional Area of Role

- **25%** Security
- **24%** IT Operations
- **22%** IAM
- **11%** DevOps / Platform engineering
- **11%** Data / ML
- **4%** Executive leadership
- **3%** Business Unit

Survey Methodology

The Cloud Security Alliance (CSA) is a not-for-profit organization with a mission to widely promote best practices and ensure cybersecurity in cloud computing and IT technologies. CSA also educates various stakeholders within these industries about security concerns in all other forms of computing. CSA's membership is a broad coalition of industry practitioners, corporations, and professional associations. One of CSA's primary goals is to conduct surveys that assess information security trends. These surveys provide information on organizations' current maturity, opinions, interests, and intentions regarding information security and technology.

Token commissioned CSA to develop a survey and report to better understand the industry's knowledge, attitudes, and opinions regarding autonomous AI agents. Token financed the project and co-developed the questionnaire with CSA research analysts. The survey was conducted online by CSA in January 2026, and it received 418 responses from IT and security professionals from organizations of various sizes and locations. CSA's research analysts performed the data analysis and interpretation for this report.

Goals of the Study

This study examines how organizations are governing AI agents as they become embedded in enterprise systems and operational workflows. The goal is to understand how autonomy is being structured, where governance practices are maturing, and where gaps remain as AI agents scale in complexity and business impact.

The survey explores several core dimensions of AI agent governance:

- **Autonomy and Control Models** – how organizations structure bounded autonomy and human intervention
- **Visibility and Discovery** – the completeness of agent inventory and exposure to shadow deployment
- **Lifecycle Management** – onboarding, permission review, ownership, and decommissioning practices
- **Policy Design and Decision Signals** – the role of risk, delegation, and contextual controls
- **Monitoring and Incident Experience** – operational oversight, incident prevalence, and business impact