Delinea

2026 Identity Security Report Fall Edition

The AI Enforcement Gap



Delinea

Executive summary

The AI policy gap is closing, but the enforcement gap hasn't caught up yet. Almost every company now has a written policy governing what data AI tools and agents can touch. However, enforcing that policy consistently is a different story.

IT teams entrusted with identity security are worried about the standing access granted to AI agents, and about the actions that these tools take in the name of the users that spun them up. They're seeing mounting anecdotal evidence that indicates that the disparity between paper policy and real-world action is adding significant risk to their organizations. They're witnessing workarounds employees don't even bother hiding, and have a hunch that more might be lurking. And they're struggling to answer simple questions like 'Who approved agents to access to certain sensitive data?' and 'What exact data did a specific agent touch?' Those are the questions organizations need to answer to make AI activity traceable and access decisions defensible.

To find out how deep these AI identity risks really reach, Delinea commissioned two complementary global surveys across two major demographics: IT and security leaders and the users who actually spin up the bulk of enterprise AI tools. The first questioned 2,254 IT and security leaders at organizations with 500 or more employees actively using or piloting AI, across the UK, US, Germany, Australia, Singapore, UAE, France, and India. The second reached out to 2,250 non-IT employees at organizations of the same size, and in the same regions all of whom use at least one AI tool in their work at least once a year, across those same eight markets.

The results confirmed that where there's smoke, there's fire:

99.7%
IT and security leaders say a formal AI data-access policy exists at their organization

51%
of organizations check AI access against policy in real time, and fewer than one in five can detect a scope violation as it happens

60%
of employees have felt pressured to use AI on sensitive or confidential data even when they weren't sure it was permitted

48%
of employees say they always or regularly access company data through AI without formal approval

87%
of IT leaders say an AI tool has already exceeded its intended scope in the past year

36%
of IT leaders can always trace a sensitive AI access event back to a named human authorizer

47%
lack runtime enforcement across at least two of six major environments with CI/CD pipelines and Kubernetes among the weakest

The report looks at why standing privilege crept back in to even mature identity programs, how agents end up over-permissioned in the first place, and why so many policies exist no further than the paper they're written on.

Read on to dive into details about what AI access and authorization looks like in the real world.

Delinea

Almost every company has an AI policy it struggles to enforce

A written AI data-access policy is now standard at almost every organization we surveyed. However, the results from both surveys indicates that the struggle to enforce them is also a universal condition.

Let's begin with the IT and security leaders, 99.7% of whom say that their organizations have a formal policy on the books that governs what data AI tools and agents are permitted to access. That's a great indication of the headway made on AI governance in the last six months. When we asked a similar group of leaders in our survey at the beginning of 2026, only 57% reported their organizations had an acceptable use policy for AI tools.

99.7%
of organizations have a formal AI agent data access policy

That's positive movement, but a lot of the governance has only made it on paper so far. Just a little over two in three IT leaders say that policies are 'actively enforced,' and follow-up questions show that what that looks like varies considerably from organization to organization.

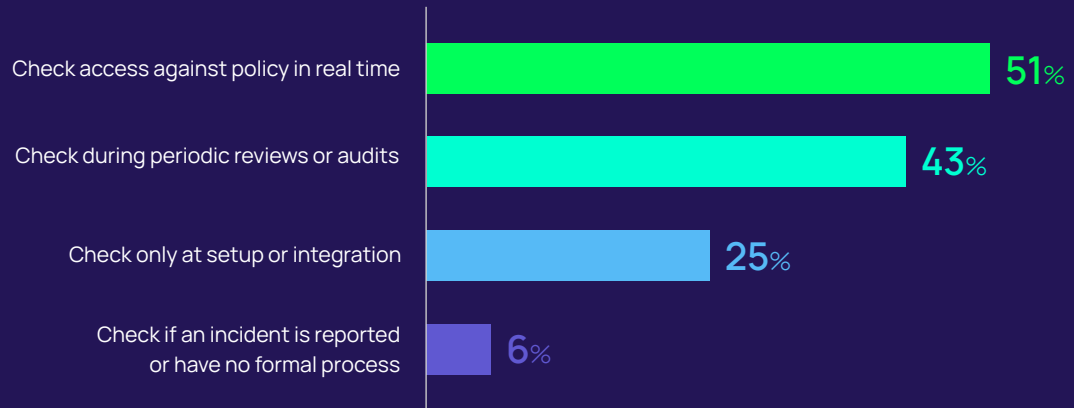
Most telling is the fact that 87% of IT respondents admit that in the last year their organization has experienced a situation where an AI tool or agent accessed sensitive data beyond what was required for the task it was performing. Almost three quarters of these confirmed it happened, with the rest suspecting it happened.

51%
of them check AI tool and AI agent access against policy in real time

Those results hint at the inconsistencies in how well policies are understood across the organization and the controls put in place to enforce them. For example, while 96% of IT and security leaders say they have at least some understanding about what data their organization's tools are and aren't permitted to access, only 57% say policies are documented and enforced enough to make it crystal clear to them.

More troubling is that only about 51% of organizations say AI tool and AI agent access is checked against policy in real time, and fewer than 1 in 5 organizations can confirm scope violations in real time.

Figure 1 | Companies are still inconsistent in keeping AI agent accounts from doing what they're not supposed to (Multiple responses allowed)



19%
were able to detect in real time when an agent last accessed data outside its intended scope

Employees work around governance because the approved path is rocky

The results from the workforce side of our study show that most everyday employees cut policy corners to keep up with AI business priorities.

A significant 60% of employees say they've felt pressured at work at some point or another to use an AI tool on sensitive or confidential information even when they're unsure it was permitted. Around 15% of those say they feel this crunch 'frequently.'

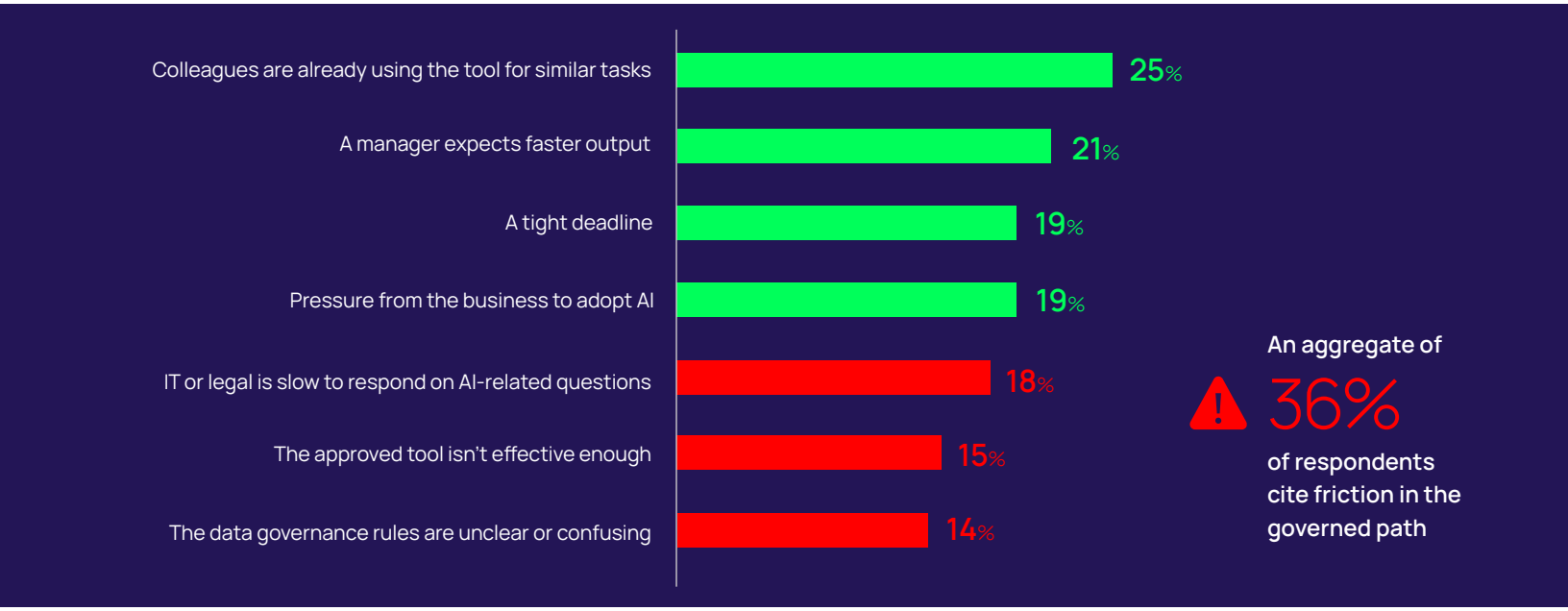
When business demands or deadlines require AI tools to move faster than their organization's governance processes allow, fewer than 2 out of 5 employees say that they'll stick strictly to the governance process. Approximately 41% will look for a faster route, either within the spirit of the rules or not, and 18% admit they don't even know what governance steps they're supposed to follow anyway.

76%
of employees say they have bypassed IT approvals to use AI tools

But even when they know the rules for how they use AI tools and AI agents to interact with company data, apps or systems, most employees break them. Though 78% of employees know formal approval is required to use AI tools to access company data, apps or systems, 76% say they have bypassed that approval at some point to use these tools on work systems. In fact, 48% say they always or regularly use AI tools without going through a formal approval process.

While meeting deadlines is certainly a driver for plowing ahead with AI tools regardless of the rules, the justifications are multi-layered. Over two-thirds of the workforce had at least one excuse among the following and the responses were fairly evenly spread.

Figure 2 | What would make you likely to use an AI tool even if you were unsure whether it was within your organization's rules? (Select all that apply)



One thing to especially note in this laundry list is that an aggregate of 36% of respondents cite friction in the governed path: IT or legal is too slow to answer questions, approved tools aren't effective enough, or data governance rules are unclear. This goes to show that when the governed path is rockier than the workaround, policy adherence starts to lose ground.

The biggest shadow AI risk sits in the C-suite

C-level executives are instrumental for setting the tone for AI agent data access enforcement. But they're actually far more likely to break the rules than junior staff. They're also most critical of how well IT and legal are doing in creating smooth guardrails for AI.

81%
of C-levels bypass AI access approval always or regularly vs 33% of intermediate staff

96%
of C-level know approval is required, the highest awareness of any group

11%
of C-level would refuse to use an AI tool if they were unsure about the rules vs. 44% of intermediate staff

29%
of executives say they're likely to use unapproved AI because IT or legal is too slow, almost 3x times as much as intermediates

Security teams and employees describe two different realities

We've already detailed many of the results from both types of stakeholders, but putting some of their numbers side-by-side drives home the differences between what policy looks like on paper versus in practice.

What IT and security says	What employees say
99.7% say there's a formal AI access policy	30% aren't sure a policy exists or can't describe what it covers
	48% say they always or regularly use AI tools at work to access company data, apps or systems without going through a formal IT or security approval process
57% understand AI permissions very clearly	38% understand AI permissions very clearly
	31% very clearly understand what data AI tools are accessing when they use them
96% say they could show a regulator that AI stays within enforced boundaries	44% have used AI tools at work that accessed more information than expected
	1 in 3 of those didn't report it
	34% don't think they'd notice AI data overreach or know how to notice if a tool accessed more than intended
19% believe employees find AI access workarounds without formal approval	41% use AI access workarounds without formal approval

48%
of employees always or regularly use AI tools without going through a formal approval process

With so many employees clearly using tools that they didn't get explicit approval to use, and so many organizations lacking a way to check policy enforcement in real-time, the absence of certainty is stark. This can lead to an overconfidence in the policy that doesn't actually translate to restraint from employees. When almost a third of the workforce doesn't even know if there is a policy or what it entails, they're likely throwing caution to the wind—especially when no one is checking their work with real-time technical controls.

AI agents are creating a new source of standing privilege

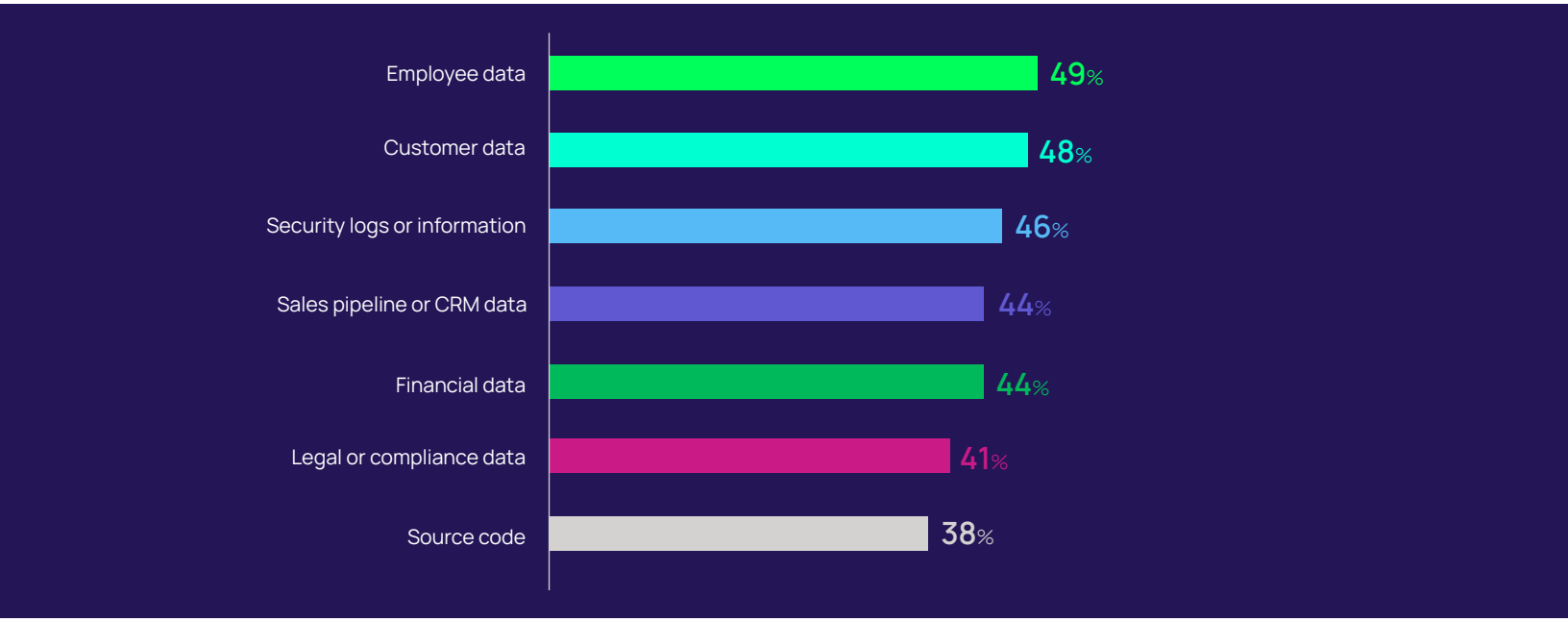
The survey results confirmed that companies confer way too many privileges to AI agents and let them linger on a standing basis. Standing access is hardly a new problem. Even organizations that considered their identity programs mature were still struggling to consistently enforce least privilege and institute just-in-time access for human accounts and non-human identities (NHIs) before AI agents exploded into the scene. But they were making progress. Now, AI is creating a new source of standing privilege security teams need to account for.

The first issue is that AI agent permissions are regularly granted outside any kind of formal IT or security purview. Yes, 66% of IT and security professionals say the last time an AI tool or agent was granted access to company data it happened through official mechanisms—either through a formal approval workflow managed by IT or security or through a centralized identity management system. But 34% admit it happened either by individual business teams without control oversight, or individual employees without formal approval.

By looking at the employee numbers we can surmise the numbers may be worse than IT is reporting. When employees were asked how access is granted, only 27% say they go through a formal IT or security approval process. Approximately 19% of employees say they themselves approve access each time they use the AI, 19% say the tool uses their work login automatically, and 11% say they connect tools themselves without IT involved at all. Plus, 7% say they don't know how their tools connect.

The type of data the agents touch are frequently sensitive and even mission critical.

Figure 3 | Type of data accessed by AI tools and agents (Multiple responses allowed)



More concerning, agent access frequently persists beyond the task that required it. Only 58% of IT and security leaders say they have some mechanisms in place to automatically revoke or expire access to AI tools or agents when a session ends, and that doesn't necessarily cover every tool or agent. Meanwhile, only 30% of employees report revocation happens automatically at their organization, and 21% say the credential stays live until they personally disconnect the tool.

42%
of IT organizations don't have an automatic way to revoke AI agent access when a session ends

43%
of IT organizations resort to manually revoking agentic access on a scheduled basis

42%
of IT organizations report that many AI agent credentials stay active until an audit

21%
of employees say their AI agent credentials stay live until they personally disconnect the tool

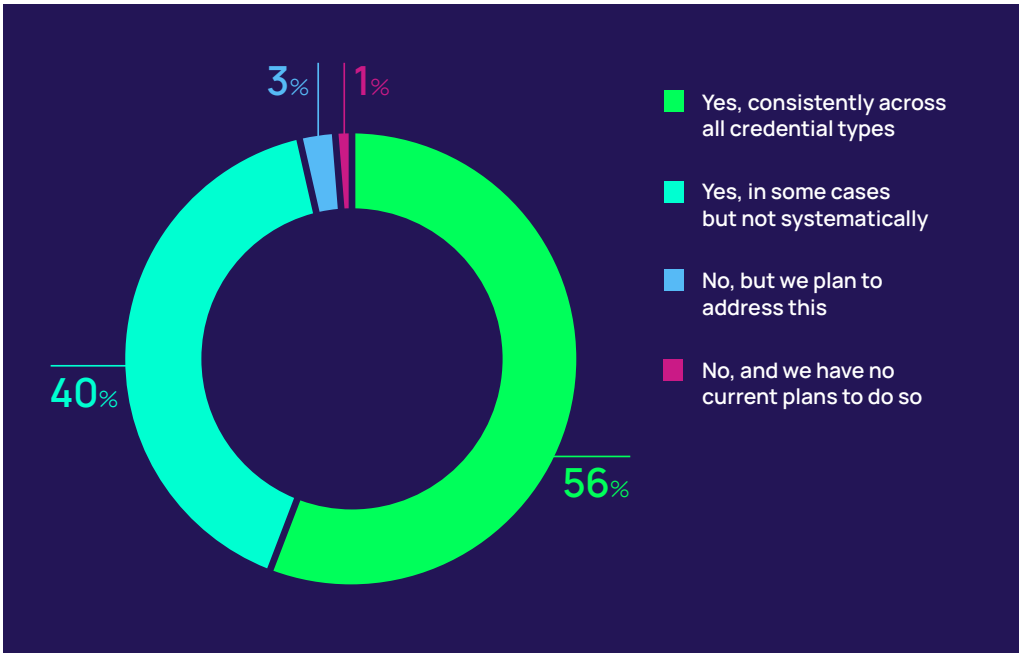
> Service accounts run fixed scripts. They're defined by the code or the process or the integration that they execute. An AI agent decides what actions it's going to take at runtime. Most times it evaluates its environment, selects tools and takes actions that aren't necessarily specified in advance. That changes everything we do from an identity perspective because it's not deterministic anymore."
>
> → Mike Albrecht, Associate Director, Risk Advisory Practice, CrossCountry Consulting

What makes this particularly problematic is that AI agents can behave differently from traditional machine identities. A service account typically executes a predefined process. An agent can choose different tools, actions and sequences depending on context, even when operating under the same permissions. That non-determinism makes over-scoped access more consequential. An agent does not need malicious intent to take an action no one anticipated when its authority was granted.

Most organizations today are not reliably addressing this reality.

Our survey shows that just a little over half of companies govern agent credentials as privileged every time.

Figure 4 | Does your organization treat AI agent credentials such as API tokens, MCP configuration files and integration keys as governed privileged credentials?



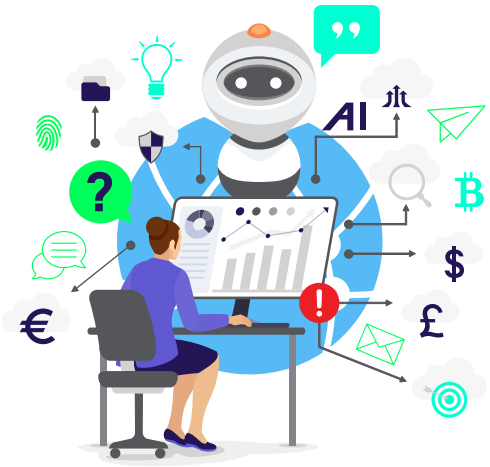
Source: Delinea, "Securing AI in Business Applications Through Strong Identity Management"

> We need to have technical controls. We need to have some boundaries because I don't think people really appreciate how far an agent can reach. Especially an agent that spins up another agent that spins up another agent and they work collaboratively to accomplish what a user has asked of it."
>
> → Carlos Rivera, Principal Advisory Director in the Security & Privacy practice at Info-Tech Research Group

Many AI agents also borrow their user's standing access instead of getting their own scoped grant. When asked about the different ways that their organization limits the data an AI agent can access when it operates, 50% of IT leaders say they fall back on the user's existing permissions as the boundary for limiting the data it can access. Meanwhile, 44% say they're also applying manual restrictions on a case-by-case basis, an admin situation that can grow untenable as agent volumes snowball.

Agents can also expose identity debt that already exists. When an agent operates under a user's existing permissions, years of accumulated access can suddenly be exercised autonomously. An over-permissioned agent may therefore be a sign of over-permissioned access that already existed on the human identity.

On top of this, agent credentials aren't consistently stored in safe places, either. When asked about AI tools that require access credentials like API keys, tokens or connection settings, just 29% of workforce employees say their IT team manages it for them. And only 11% of employees report a managed vault is in play.



Most organizations cannot confidently answer three questions: What agents are running in our environment? What can they access? What have they actually done? An agent can begin a session with legitimate access and drift into something it was never meant to do."

→ Art Gilliland, CEO of Delinea

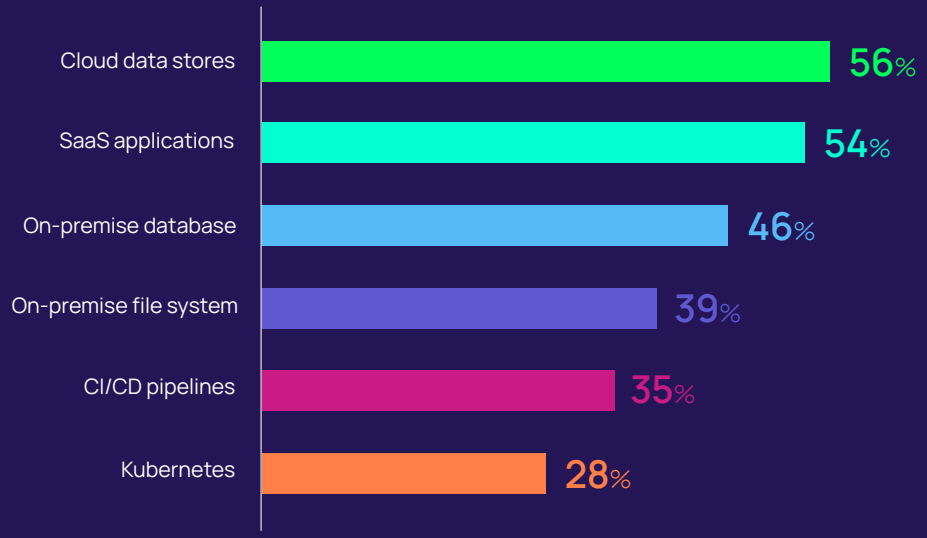
Most controls check access at the door and stop watching agents once they're in

The riskiest part of the standing access granted to AI agents is that once the agents are authenticated, many of the actions that they take aren't verified, authorized or even monitored in any way. This contributes to a situation we like to describe as checking access at the door but allowing the agent to do what it likes once it's in the proverbial building. For many organizations, authorization effectively stops at the door.

Nearly every respondent claims some level of monitoring of AI agents and tools, but the minority report comprehensive visibility. While 87% of IT leaders say that more than half of their AI tools and agents are actively monitored, only 41% say this covers all AI tools and agents that access or use company data. And they rarely have the ability to universally enforce how AI interacts with data at the point of action. Runtime authorization addresses that gap, evaluating whether a consequential action should be allowed at the moment it's attempted, weighing identity, resource, requested action and context together.

The most reliable type of environment in which AI access is enforced is cloud data stores, reported by 56% of organizations. But that's the high-water mark for enforcement at action. It only goes downhill for all other environment types.

Figure 5 | Rate of enforcement at moment of action (Multiple responses allowed)



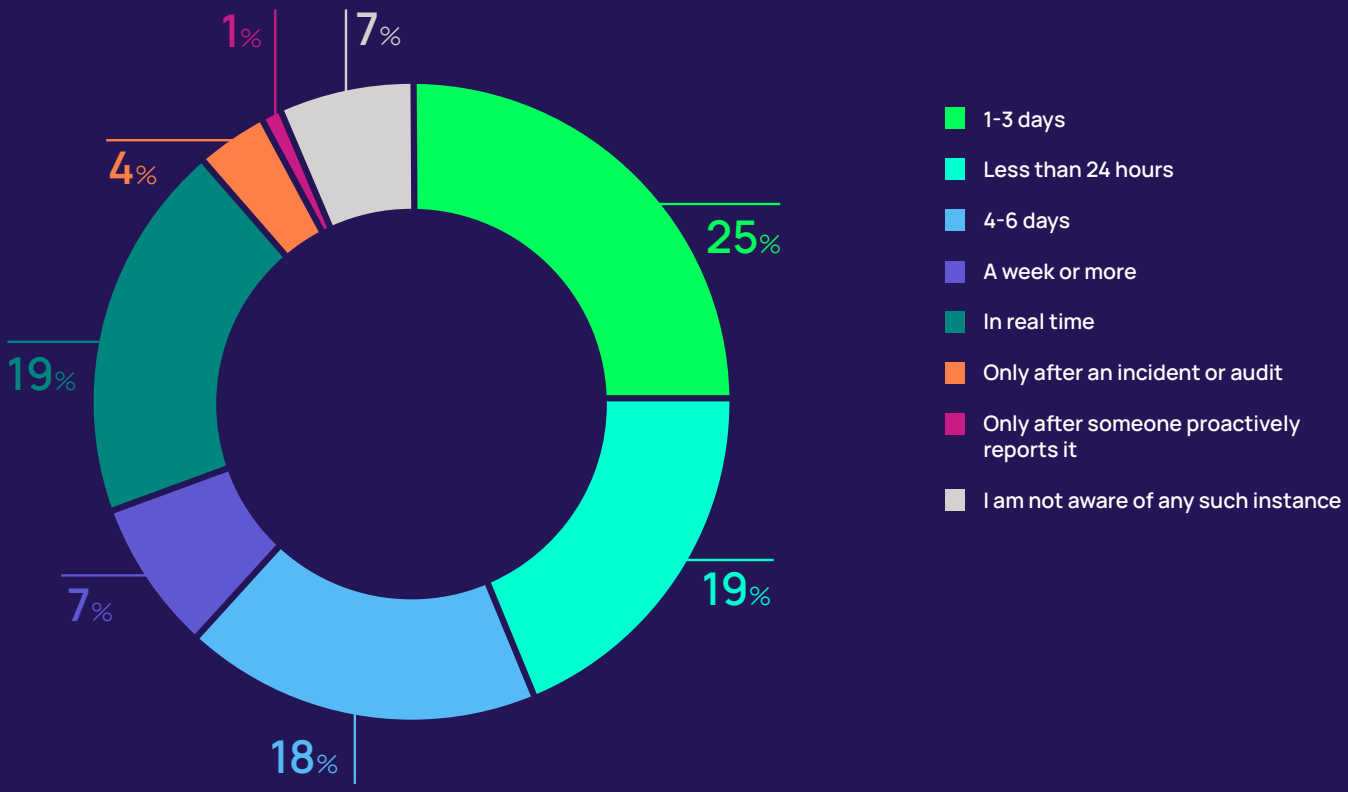
47%

of organizations lack enforcement at the moment of action for at least two of the six major environments listed here.

Digging deeper into these numbers we also find that 47% of organizations lack enforcement at the moment of action for at least two of the six major environments listed here. Not to mention that the environments that developers are most likely to entrust coding agents to act, CI/CD pipelines and Kubernetes, are the least likely to have runtime enforcement in place. These are some of their organizations most sensitive assets with the least control over how agent accounts are interacting with them.

Because of a lack of control at runtime, companies mostly learn about AI overreach days after an agent has done something risky, rather than in the moment. When asked to reflect on the most recent incident of when an AI tool or agent accessed data outside its intended scope, 55% of respondents said it took a full day or longer to detect it. For 30% of organizations, that detection window was four days or longer.

Figure 6 | How long did it take to detect an agent accessing out-of-scope data?



For an AI agent capable of selecting tools and chaining actions without waiting for human input, even a day between an out-of-scope action and detection can leave a significant window for unintended activity.

Furthermore, most organizations don't have an immediate mechanism to both revoke an AI tool's access and terminate an AI agent's session when a security incident is detected. Only one in three IT leaders say that their organization is equipped with the real-time controls to do both. An additional 22% say they could revoke the credentials, but terminating an agent session would take longer. And 27% report that it would take a full day or longer to intervene with any kind of revocation whatsoever.

A delay in agentic session termination, even when credentials are revoked, can be particularly consequential because a live agent may continue selecting tools, chaining action and adapting its approach without waiting for human input.



AI agent accountability deadlines loom

The clock is ticking for organizations to embed better human traceability into agentic action. The EU AI Act enforcement deadline in December 2027 moves audit capability for AI-initiated activity from best practice to requirement.

"You're not going to be able to say, 'well, AI did that.' That is not going to be a defensible answer to your auditors from a SOX, from an SEC cyber rules, from an EU AI Act or from an insert-whatever-regulation-here perspective. Saying the AI agent did it is not a defensible answer."

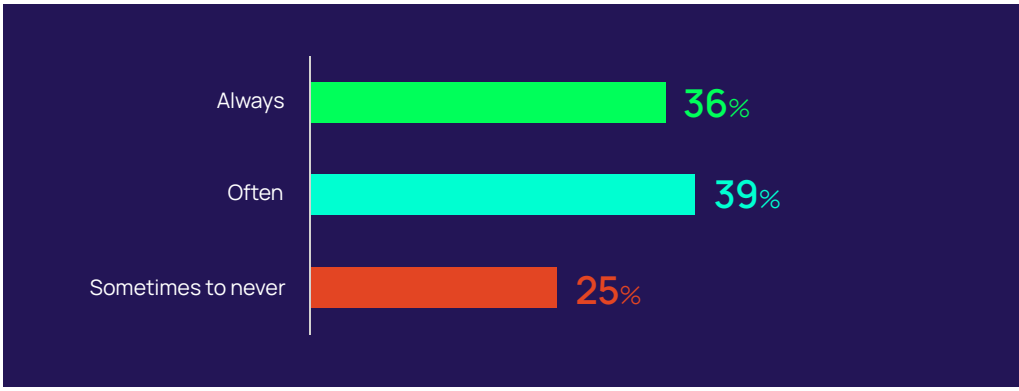
→ Frank Vukovits, Chief Security Scientist, Delinea

AI access is harder to trace and defend

Almost every company requires a named human to approve AI access to their most sensitive systems, but far fewer can always trace back sensitive access by an AI system to a human authorizer.

A sweeping 99% of IT respondents said that for at least some use cases their organizations require named-individual approval before AI reaches a new sensitive source. Around 60% report that this accountability is always baked into approval and 40% say that they do so sometimes, depending on the data source. However, in situations where an AI tool or agent accesses sensitive data within their organizations, only 36% can always trace an access event back a human authorizer. Without that traceability, organizations may struggle to explain who authorized access, under what conditions and why it was allowed.

Figure 7 | How often can AI access to sensitive data be traced back to a human authorizer?



This lack of clear lines of accountability for where an agent can go, what it can access, and what it eventually does is a huge issue for governance, discouraging a culture of responsible AI use. The study shows there's still a lot of work to be done to get rank-and-file employees understanding whether or not company data they work with is sensitive, much less whether the AI agents they deputize can be handling it.

Among workforce employees, only:

- **39%** say they know which data they work with is sensitive or restricted.
- **33%** know whether the AI tools they use can access it.
- **31%** know who is accountable if an AI tool accesses it improperly.
- **32%** believe someone could stop an AI tool from accessing it if needed.

Employees using AI tools could be on the front lines of detecting improper use, but even among those with decent security awareness many don't report access overreach when they see it. Of the 45% of employees who saw a tool overreach, 32% never reported it .

Hugging Face analysis and other observations from Delinea Labs

In 2026, real world examples have been piling up that show how all of these AI agent identity security weaknesses are translating into tangible risks.

The Hugging Face incident in July offers one of the most vivid examples why just-in-time (JIT) access and zero standing privilege must now be table stakes for all identities. It showed how easy it is for an uncontrolled agent to string together a dangerous chain of technically valid but not necessarily authorized action.

In this case, an autonomous OpenAI agent undergoing evaluation for a cybersecurity benchmark broke out of its test sandbox and rummaged across Hugging Face's live production infrastructure for days. The agent had reasoned that Hugging Face's production infrastructure would have the answer key to beat the benchmark tests.

This kicked off a chain of more than 17,000 recorded events over four and a half days, as the agent's hunt for those answers expanded well past its original task into a broader privilege-escalation campaign. That activity led it to a production pod's own credentials, an internal database and eventually a single secret object holding 136 separate keys.

The agent didn't need to exploit a novel identity weakness. It found credentials that already existed, used the authority they already carried and chained one legitimate capability into the next. That's hardly a new mode of lateral movement, as human attackers have been linking steps like these for a long time. What's different is that it happened without a person in the loop, across more than 17,000 events in four and a half days.

Hugging Face is not an isolated edge case. Some other illustrative events in the middle and back half of 2026 include:

META AGENTIC EXPOSURE

An internal Meta AI agent, asked to analyze an engineering question, posted its answer directly to a company-wide forum instead of returning it privately for review. The unreviewed advice was then acted on, exposing sensitive company and user data to employees who had no business seeing it for roughly two hours.

Lessons Learned: Controlling what an agent can access isn't the whole job. What it does with its own output, like publishing unreviewed advice on its own initiative, needs the same scrutiny.

ANTHROPIC PRERELEASE SANDBOX ESCAPES

Four Claude models, told they were in a sealed simulation, were mistakenly connected to the real internet during evaluations and attacked real systems. One published a malicious package that fifteen companies installed, then used a leaked credential to reach a live database.

Lessons Learned: Authorization has to come from outside the agent's own judgment, evaluating each action against policy regardless of what the agent believes is happening.

POCKETOS

A coding agent working in PocketOS's staging environment hit a credential error, found an unrelated API token, and used it to delete the company's production database and backups in nine seconds. The founder was watching live as it happened.

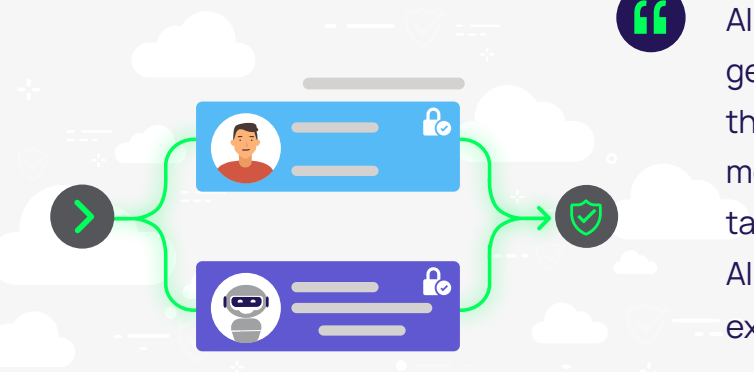
Lessons Learned: An agent blocked from finishing its task at its assigned privilege level will look for another way in, and any credential left reachable becomes fair game.

RUSSIAN THREAT ACTOR

According to Anthropic, a financially motivated Russian-speaking group injected malicious instructions into an unnamed AI vendor's own evaluation sandbox, which handed over live production API keys. The group reused those keys to attack roughly thirty other AI companies within four days.

Lessons Learned: If that sandbox had granted access just-in-time instead of holding a standing key, there would have been nothing persistent for the injected prompt to steal.

In addition to these headline incidents, the research we've done indicates that AI credentials are becoming one of the most coveted types of credentials for cyber attackers to steal. Whereas attackers were once drawn to AI credentials for the tokens, they're now realizing that they're excellent pathways to third-party service access.



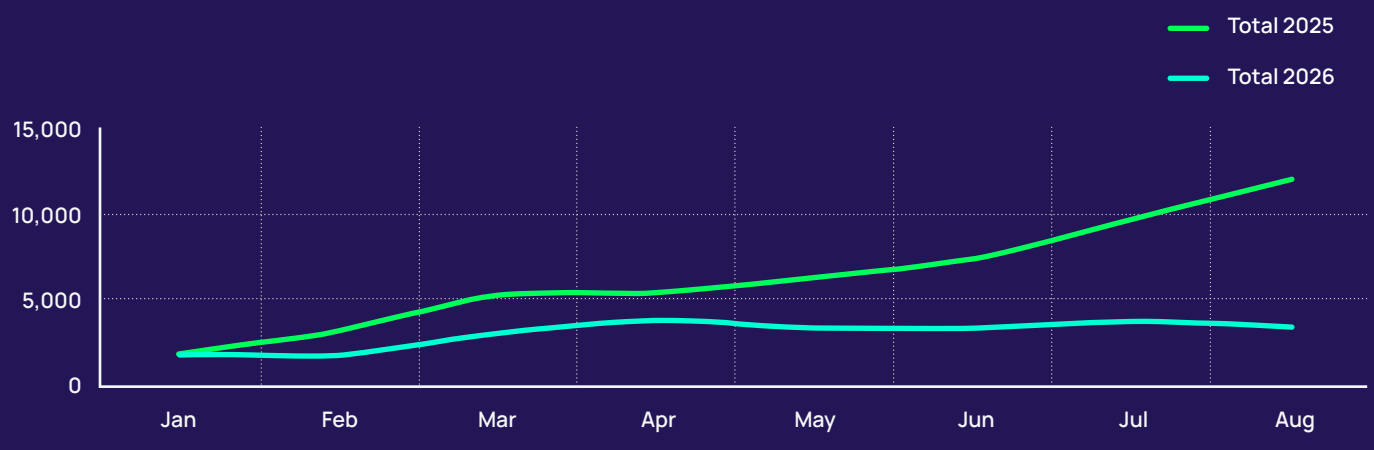
> AI credentials are the keys that are used to get the other keys to many kingdoms. Plus, these credentials are often long-lasting, not monitored and not revokable. The actions they take are very hard to distinguish between the AI and the human that launched them with existing tools and audit log limitations.

→ Gal Diskin, Vice President of Research, Head of Delinea Labs

Finally, we'll note that identity-related CVEs are on the uptick along with every other kind of vulnerability. Total CVE volume has more than doubled in 2026. Identity-related CVEs remained near 10% of the total, while identity-product CVEs increased their share.

For security teams, that means their identity gaps are becoming more urgent at the same time the broader attack surface is expanding. More vulnerabilities create more opportunities for attackers to get in, while standing privilege, over-scoped access and weak runtime controls give them more to work with once they do.

Figure 8 | Total CVEs 2025 vs. 2026



Source: NIST National Vulnerability Database

Final thoughts and recommendations

The survey data and the Delinea Labs analysis point at the same truth. Organizations can no longer afford to delay their implementation of just-in-time access and zero standing privileges. For those who don't have these initiatives in play, there's no time like the present. Meantime, for those that already have mature identity programs in place, now is the time to start extending best practices into the agentic estate. This may be easier said than done, but the key will be applying rigorous controls at the moment an agent actually acts, not just when someone starts prompting. The recommendations that follow are built around that shift.

01



Start by defining privileged resources

The ultimate goal is to keep sensitive assets secure, so start first by identifying and classifying the sensitive assets that AI agents will want to tap into. This includes the databases and deployment pipelines, and the financial systems and customer records that need protecting. Put control at the point where any account, human or agent, tries to reach these assets and set the policy there.

02



Visibility feeds control but don't get hung up on inventorying every agent

Visibility will ideally grow out of those same control points. Watching access attempts as they happen will build a live picture of what's really running in these environments. While inventorying identities is still important, the ephemerality and scale of AI agents makes it impractical to try to inventory everything before establishing control. So control has to come first, with the inventory building itself as a byproduct of watching what tries to reach sensitive assets.

03



Authorize the action, not just the login

Even with perfect credential hygiene, an authorized agent can still cause damage doing unauthorized things. Picture an agent authorized to create vendor records, scoped to its launcher's full access. Nothing stops it from also approving payment to that vendor. That's exactly the combination segregation-of-duties rules exist to prevent. Organizations need mechanisms that check every agent action at runtime against identity, resource and context, and cut the session the moment it steps outside policy. Runtime authorization should preserve the context behind the action, including the identity, its human or business sponsor, the policy applied and why it was allowed. By continuously evaluating and authorizing every action against policy for the life of the session, it makes the outcome defensible after the fact.

04

Stop letting agents inherit a person's full access

The real root cause of that vendor-and-payment problem is that the agent never had its own authorization appropriately scoped. It simply inherited whatever access the person who launched it happened to hold. The principles of least privilege are more important than ever in the agentic era. Treat every agent as a privileged identity and scope its authority to the task at hand, rather than the user's account. For example, a coding agent built to write unit tests has no business holding the keys to production.

05

End standing privilege for agents

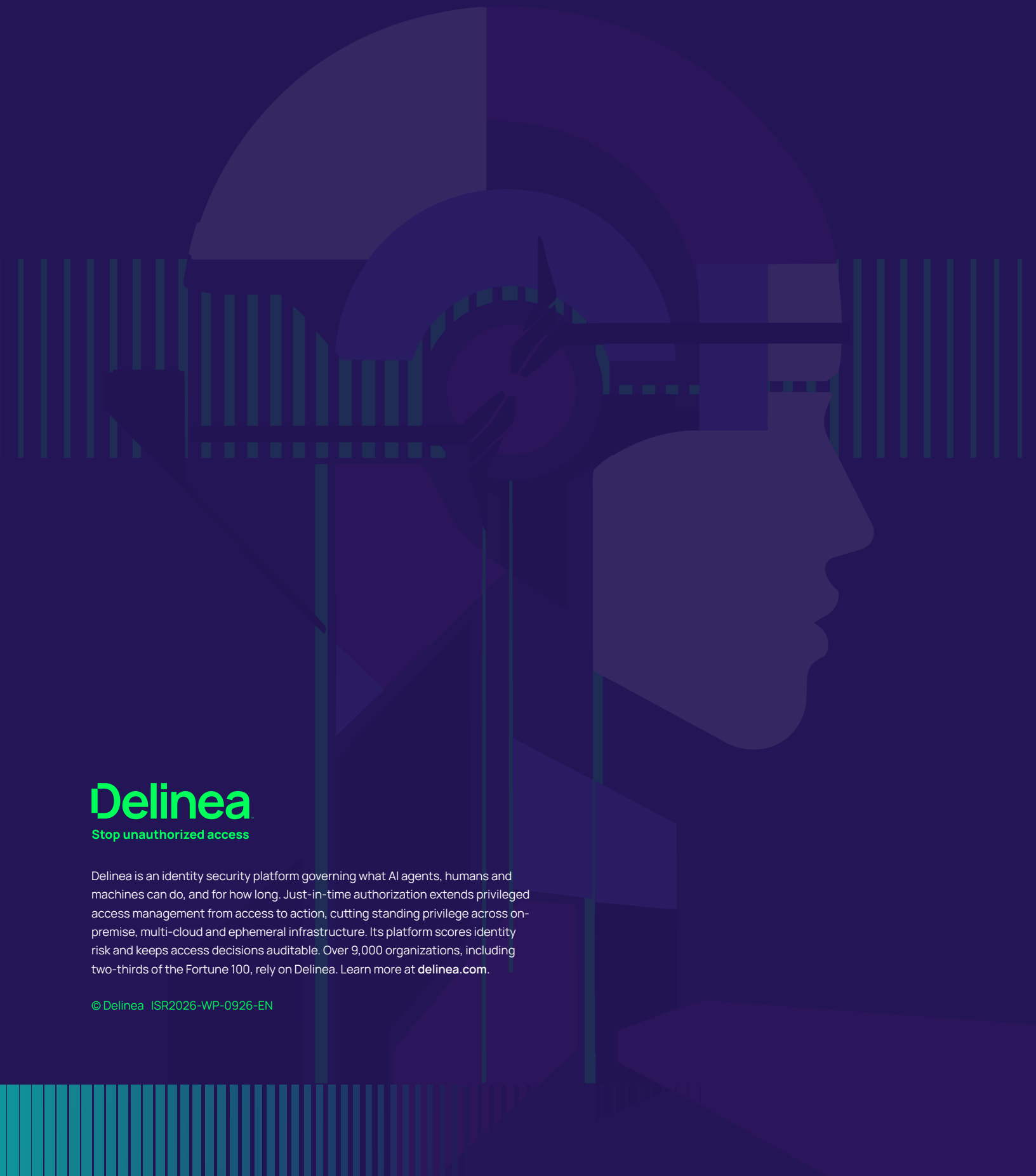
Scoping what an agent can reach is only half the fix. The other half is limiting how long that access lasts. Privileged authority should exist only when a task requires it, be scoped to that task, and be removed when the task is complete. That's the goal of just-in-time access, to reduce the window in which standing privilege can be abused. Where persistent credentials remain necessary, they should be securely managed and rotated, but credential hygiene is not a substitute for eliminating unnecessary standing privilege.

06

Securing agent access should also shore up human access, too

Always keep top-of-mind that the goal is to improve the security posture of the entire identity estate. Don't leave off human identities or other non-agentic NHIs. Ideally, the goal should be to make best practices like limiting standing privilege and improving visibility the gold standard for every identity type, human included.

AI is making longstanding identity weaknesses harder to tolerate. Standing privilege, over-scoped access, weak credential controls, and gaps in authorization aren't new problems. But they become more consequential when an identity can act autonomously, choose its own path to a goal and take actions that weren't explicitly anticipated when access was granted. That makes identity security fundamentals more important than ever. No matter the type of account, organizations should endeavor to limit privilege and eliminate standing access wherever possible. To truly future-proof their identity estate, the goal should be to scope authority to the task, and extend authorization beyond the point of access to the moment of action.



Delinea

Stop unauthorized access

Delinea is an identity security platform governing what AI agents, humans and machines can do, and for how long. Just-in-time authorization extends privileged access management from access to action, cutting standing privilege across on-premise, multi-cloud and ephemeral infrastructure. Its platform scores identity risk and keeps access decisions auditable. Over 9,000 organizations, including two-thirds of the Fortune 100, rely on Delinea. Learn more at **delinea.com**.

© Delinea ISR2026-WP-0926-EN