Delinea

Uncovering the Hidden Risks of the AI Race

Delinea

Contents

Introduction

The AI mandate from business leaders is clear:
To remain competitive, you must accelerate AI adoption.

Operationally, organizations can't afford to let security friction hang up agentic AI deployments. Executive teams are under pressure to innovate quickly and keep pace with competitors. Even everyday non-tech workers are building AI skills to stay ahead of their peers and maintain their relevance in the workforce. As a result, teams across the business are moving quickly to deploy agentic AI.

Unfortunately, legacy security models built for humans aren't evolving fast enough to fully monitor new agentic AI operating models. In the rush to innovate, identity controls are often relaxed, inconsistently applied, left undefined, or ignored entirely.

To understand how organizations are navigating these challenges, Delinea commissioned Censuswide to run a global survey of 2,001 IT decision-makers who are actively using or piloting AI in their environments across the UK, US, Germany, France, Australia, Singapore, and India.

The study finds that in moving to agentic AI operating models, **organizations are introducing substantial identity-related risks**—much of which remains outside the scope of traditional visibility and governance processes.

We surveyed those actively using or piloting AI

2,001
IT decision-makers

7
countries

Delinea

The findings reveal that organizations express high confidence in their security readiness for AI while simultaneously admitting they lack the fundamentals to back up that confidence.

They acknowledge gaps in identity discovery, monitoring, and privilege control. Under the strain of the opportunity risks of falling behind in the AI race, risk managers are under constant pressure to loosen identity controls—and they often do.

Clearly, organizations can't afford to slow down AI adoption. But the study indicates that identity security must evolve alongside AI adoption. Leaders must modernize the way they discover and protect access and identity relationships in the AI era so they can innovate quickly without abandoning identity governance in the process.
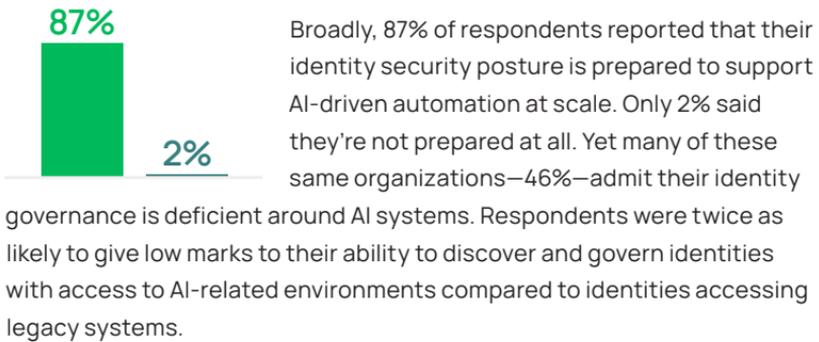
2.0

The AI security confidence paradox

Delinea

The survey findings reveal a clear gap in agentic AI preparedness. Throughout the research, respondents held two conflicting beliefs: Most are highly confident that their systems are ready for agentic AI at scale, but many also acknowledge gaps in their ability to govern AI-related identities.

This dynamic exposes what we call the "AI security confidence paradox."

87%

2%

Broadly, 87% of respondents reported that their identity security posture is prepared to support AI-driven automation at scale. Only 2% said they're not prepared at all. Yet many of these same organizations—46%—admit their identity governance is deficient around AI systems. Respondents were twice as likely to give low marks to their ability to discover and govern identities with access to AI-related environments compared to identities accessing legacy systems.

2x **Respondents were twice as likely to give low marks to their ability to discover and govern identities with access to AI-related environments compared to identities accessing legacy systems.**

The pattern held across multiple lines of questioning. Confidence in discovery and protection of identities in AI environments was consistently high. However, follow-up responses uncovered limited validation mechanisms and incomplete oversight.

For example, while 82% of organizations said they're very confident in their ability to discover non-human identities (NHIs) with access to production systems. Fewer than 1 in 3 organizations actually validate NHI and AI agent inventory usage or access patterns in real-time to ensure discovery is working.

This paradoxical thinking indicates organizations may be advancing agentic AI without fully modernizing the identity controls required to support it. They may not yet realize the level of risk incurred by agentic AI. As the rest of the survey results show, this is likely because their beliefs are built around incomplete information.

Confidence vs. real-time validation

82%

Very confident in NHI discovery

30%

Validate NHI/AI usage in real time
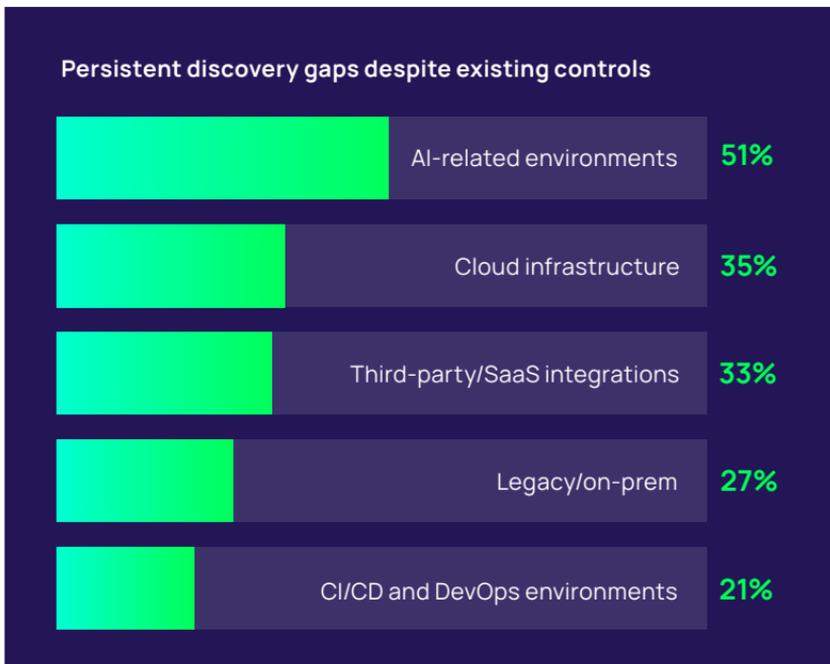
Delinea

3.0

The identity visibility gap: What you don't know can hurt you

Our data shows that most organizations haven't yet built the mechanisms to fully show them how identities are used, whether by humans or AI agents.

90%

An overwhelming amount of respondents admit to having at least some sort of identity visibility gap at their organization.

The number one gap was machine and NHI accounts, including those used by AI agents. Respondents reported that the **identity discovery gaps most likely to persist over time were in AI-related environments, at nearly double the rate of legacy** and on-premises systems:

Persistent discovery gaps despite existing controls

AI-related environments	**51%**
Cloud infrastructure	**35%**
Third-party/SaaS integrations	**33%**
Legacy/on-prem	**27%**
CI/CD and DevOps environments	**21%**

The challenge is that awareness doesn't always translate into corrective action. Many organizations have not yet experienced a security incident linked to AI-related identity weaknesses. Without a measurable failure or breach tied directly to AI identity blind spots, the risk remains an abstract rather than an operational certainty.

Delinea

42% say AI expansion is one of the top factors that has increased their NHI risk in the past 12 months.

Most respondents acknowledge they have worries about AI agent access. Approximately 42% of organizations admit that AI expansion is one of the top factors that has increased their NHI risk in the past 12 months—far more than increased automation and CI/CD velocity (26%) or growth in cloud native workloads (26%). Fewer than 1 in 10 said that there's no particular risk that worries them about NHIs and AI agents. Meanwhile, 38% said they're worried most about excessive autonomy or privilege, 35% said they are concerned about limited auditability and explainability, and 32% said they're worried about rapid identity proliferation due to these accounts.

Until they close the visibility gap, these concerns remain vague worries while the threats accumulate unseen in their environments. Without visibility into NHI and AI agent activity, organizations can't detect anomalous behavior or investigate suspicious actions.

Even more detrimentally, the lack of visibility into machine identities also leads to uninformed risk acceptance.

> **The business is accepting AI risk to stay competitive, but because AI is such a new paradigm, they're accepting it without actually understanding qualitatively or quantitatively what the risk is. It's so new that CISOs can't yet convey 'here's the risk' in concrete terms. That's a gap on the GRC professional's side, too."**
>
> Dr. Gerald Auger, Ph.D, head of Simply Cyber

Identity type currently creates the largest visibility gap in your organization today

Machine/non-human identities
(bots, AI agents, service accounts) — **40%**

General workforce (employees, contractors) — **36%**

The fact that NHIs only narrowly beat out workforce identities speaks to another, deeper issue. Organizations haven't yet perfected identity visibility governance for human users before machines started to proliferate, so AI is exacerbating an issue that has been lingering for a long time.

Delinea
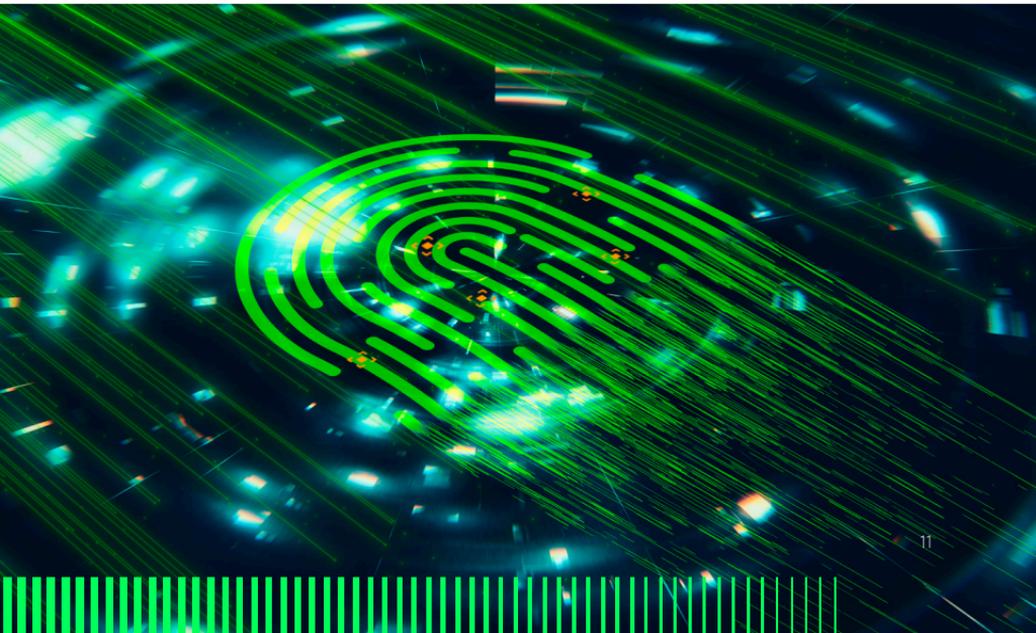
4.0

Why identity weaknesses in AI remain invisible

Our survey analysis found three major areas that contribute to the systemic lack of visibility into AI-related identity risks:

▶ **Speed prioritized over governance:** Innovators accelerate without fully integrating identity governance controls, leading to exceptions, inconsistent policy enforcement, and unmanaged access.

▶ **Expansion of shadow AI:** Widespread use of unsanctioned or independently deployed AI creates identities and access paths that fall outside established discovery and protection processes.

▶ **AI fueling unchecked NHI activity:** Using machine and AI identities that are given standing privileged access without monitoring, validation, or least-privilege enforcement is growing.

These factors are interconnected, with each presenting its own set of challenges and consequences.

Speed prioritized over governance

The operational and opportunity risks of slowing down AI innovation are outweighing the security risks in many business decision-makers' minds. Many CEOs see an inability to keep up with AI as an existential business problem. With developments advancing rapidly, they're directing every part of the business to get rid of the obstacles keeping them from maximizing value from AI advancement.

Unfortunately, identity and access control can cause friction in business systems, IT environments, or developer pipelines. Respondents to the survey noted moderate to high friction in every kind of workflow, from cloud provisioning to third-party vendor access.

Delinea

Workflow friction attributed to identity and access controls

Joiner-Mover-Leaver (JML) processes		
Moderate	39%	**65%**
High	26%	
Privileged operational access (incident response, break-glass)		
Moderate	40%	**71%**
High	31%	
Deploying AI agents		
Moderate	36%	**74%**
High	38%	
Automation workflows		
Moderate	35%	**75%**
High	40%	
CI/CD pipelines		
Moderate	41%	**75%**
High	24%	
Cloud and infrastructure provisioning		
Moderate	35%	**79%**
High	44%	
Third-party or vendor access		
Moderate	43%	**75%**
High	32%	

Friction around agentic AI deployment and automation was particularly high, alongside CI/CD pipelines and cloud provisioning, which tend to be highly automated and ephemeral in modern IT operating models.

The friction identified in the survey, along with identity misconfiguration, has measurable business impacts:

39%
of respondents reported increased operational cost

37%
reported increased operational complexity due to fragmented identity tools

34%
reported delayed releases or delayed AI initiatives

As organizations weigh the friction introduced by legacy identity controls against the perceived urgency of AI innovation, governance standards are often relaxed to maintain deployment velocity. Survey participants reported significant business and engineering pressure to reduce privileged access controls to ensure AI-driven automation "just works." Almost all organizations—90%—place pressure on security teams to loosen access control to support AI-driven automation, with nearly 1 in 5 organizations noting that it is strong pressure.

Pressure on security teams to loosen access control for AI

| 20% | 90% | |

Strong Pressure

Delinea

When security requirements conflict with business speed, fewer than 1 in 3 organizations say that security requirements are consistently enforced.

Approximately 11% of respondents say controls are bypassed altogether by shadow use, a topic discussed next. A quarter of respondents report that exceptions are granted on a case-by-case basis. Another quarter says controls are either temporarily disabled or standing privileges are granted. The last two are lumped together because, in practice, security veterans will tell you that when it comes to granting access, "temporarily disabled" rarely gets revisited.

"Let me tell you from doing security in the real world, they will never be re-enabled. That's typically how these things play out," Auger says.

The pattern of prioritizing speed over governance is not new to security leaders. The risk management community has seen this dynamic play out with bring your own device situations, cloud adoption, SaaS sprawl, and now agentic AI use.

> **We keep saying we need to build security in, not bolt it on. But then, every new tech paradigm we give a security hall pass. People go out and do a bunch of innovative new greenfield projects, and security has to come in after the fact and try to harden it."**
>
> Chris Hughes, Resilient Cyber

Experts say that because of the friction problem, identity tends to be the element most consistently left behind when governance is loosened in favor of speed.

> **❝** **We've gone through these technological advancements and transformation programs time and time again, but we never seem to recognize the correlation. There's a common thread that pulls across all of them, which is identity."**
>
> Kayla Williams, vCISO and SANS Institute's 2024 CISO of the year

The pressure comes from the very legitimate fear of falling behind competitors who choose to move faster: "It's being driven by the fact that the business feels like they have to keep pace. Our competitors are doing this. Or at the user level, our peers are doing this. If we don't, we're going to fall behind," Hughes says.

The goal should be to stay out of cynical defeatism and try to learn from past cycles. Hughes believes that security teams can adopt the same AI technologies and innovations to improve their own operations and move fast enough to work at the pace of agentic AI innovation.

Rampant shadow AI

Industry reports have resoundingly sounded the alarm on shadow AI, with analysts at Gartner estimating that **by 2030, some 40% of organizations will suffer security incidents due to shadow AI risks.**[1] Meanwhile, recent employee surveys are showing that **80% of workers admit they've used unapproved AI tools in their jobs.**[2]

The rapid, runaway adoption of OpenClaw in early 2026 indicates that the use cases for these unapproved apps are growing increasingly risky. Users aren't just doing quick spell checks with Grammarly or asking ChatGPT simple questions. They're running full-fledged agents with sweeping permissions.

---

[1] https://www.infosecurity-magazine.com/news/gartner-40-firms-hit-shadow-ai/
[2] https://content.upguard.com/hubfs/resources/The-State-Of-Shadow-AI-Report-2025.pdf

Delinea

As an open-source AI assistant that can modify files and execute commands without any intervention, OpenClaw has persistent memory and broad permissions. It's typically granted direct connections to the networks and enterprise services used by the machine it's installed on.

In late January, security researchers at Censys tracked a steep adoption curve that saw publicly exposed OpenClaw instances grow from 1,000 to 21,000 instances in a single week.[3] More concerning, a report from TrendMicro found that during this ramp-up, 1 in 5 organizations had employees deploy OpenClaw without IT approval.[4]

"What ChatGPT did by putting a text field in front of the LLM and unlocking it for normal people, OpenClaw is doing for agents," explains Auger. "Carl, who works in accounting and doesn't have a technical bone in his body, can just copy and paste a command. And all of a sudden, he's got an agent running on his machine under his permissions. And now Carl's like, 'This is sick. I'm just going to watch YouTube while this thing does my job.'"

This is a glaring example of the shadow AI problem highlighted by our study. **A significant 53% of respondents said they're regularly encountering unsanctioned AI tools and agents accessing company systems or data.** And that's just the ones they're detecting.

While 79% of respondents report some level of confidence in detecting unsanctioned AI tools or agents, many respondents hedged their answers slightly—with just 37% reporting they were very confident. Respondents say that most detection takes at least hours to days to occur, with only 28% reporting the ability to detect shadow AI in real-time.

[3] https://censys.com/blog/openclaw-in-the-wild-mapping-the-public-exposure-of-a-viral-ai-assistant
[4] https://www.trendmicro.com/en_us/research/26/b/what-openclaw-reveals-about-agentic-assistants.html

Shadow AI is already inside

53%

regularly encounter
unsanctioned AI tools
or agents accessing
company systems

79%
have some confidence
in detecting shadow AI

28%
can detect shadow
AI in real-time

The struggle to detect access granted to unauthorized AI agents can
add tremendous risk to organizations because shadow AI is so difficult to
distinguish from normal activity.

"Agents behave exactly like compromised credentials: They're trusted,
they're persistent, they're pretty much invisible once they're inside,"
Williams says.

What makes the current wave of shadow AI particularly thorny to address is
the grassroots nature of adoption. As Hughes explains, **"It's a bottom-up
issue. It's developer-led but also everyday-employee-led. Everyone sees
it as a productivity boost."**

It isn't just executive leadership teams who worry about getting left behind.
Everyday employees recognize that AI is reshaping how work gets done and
are adopting AI tools to remain relevant in their roles. Many are deploying
AI agents to do their jobs, effectively becoming citizen developers in
the process.

Delinea

"We're facing an entire workforce who are becoming 'coders,' and many of them are not technically savvy enough to understand sandboxing or at least permissive access control," Hughes says. "Anyone can use these tools and grant broad access and permissions without really understanding the implications."

This creates decentralized deployment with centralized risk, creating significant visibility and governance challenges for risk management teams.

"Anyone can deploy it super easily, which introduces major risk," Auger says. "Not only is it sprawling—it's also getting all these permissions. And if you give it your own credentials, I can't tell on the network if it's Carl or Carl's AI."

AI fueling unchecked NHI activity

NHIs were an identity governance challenge long before agentic AI took the scene. IoT devices, microservices, API connectors, and automated software engineering processes all require machine accounts with varying degrees of privilege. Trying to gain visibility into NHI access was already overwhelming the legacy identity security processes rooted in human-centric identity controls. **Two years ago, analysts estimated NHIs outnumbered human accounts 46 to 1. A year ago, the industry estimate almost doubled to 82 to 1.**[5]

Two years ago

One year ago

Rapid adoption of agentic AI expands the number of non-human identities (NHIs) within the identity estate and introduces new layers of operational risk. AI agent accounts don't follow the pre-programmed, deterministic steps characteristic of previous generations of automation. AI Agents make contextual decisions and initiate actions that weren't explicitly scripted. As a result, **they may request additional access, interact with new systems, or trigger privilege changes dynamically, essentially escalating their own privileges.** Without continuous discovery, validation, and least-privilege enforcement, this behavior can create unmanaged access paths and unintended privilege exposure.

"The biggest difference is that with other generations of automation, we knew exactly what it was designed to do," Hughes says. "Agentic AI does things we didn't anticipate because it's non-deterministic. And in many cases, it can be hard to even understand why it did something or why it asked for elevated privileges to do that action."

80%

Our survey results highlight this challenge: It found that 80% of organizations today are unable to always understand why an NHI took a privileged action.

Agentic AI will continue to supercharge the NHI issue as organizations scramble to get AI agents operational without running into authorization issues. These accounts are frequently accessing and connecting to business-critical systems with very little governance in place. According to Cloud Security Alliance's 2026 State of NHI and AI Security report, less than a quarter of organizations today have documented and formally adopted policies for creating or removing AI identities.[6]

---

[6] https://cloudsecurityalliance.org/artifacts/state-of-nhi-and-ai-security-survey-report

Delinea

AI capability = AI risk

"The reason these tools are so vulnerable and potentially risky is because access gives them so much utility."

Chris Hughes, Resilient Cyber

The very things that make agentic AI valuable are what make it risky: broad access and connectivity across systems. AI agents derive their utility from their ability to tap into multiple systems, pull data from various sources, and take autonomous action based on what they find.

But that same access and connectivity is exactly what creates risk exposure.

For the sake of speed, these accounts are often given far more privileges than the typical worker. Our study shows that most organizations are granting always-on or standing access privileges to NHIs and AI agents to keep barreling forward with adoption, even if they're aware that it significantly increases their risk posture.

The survey found:

73%

of respondents agree that standing access for NHI and AI agents increases risk

74%

say standing access for NHIs and AI agents is necessary to meet uptime expectations

68%

say security teams often accept standing access for NHIs and AI agents under operational pressure

"If I, as an employee, had this level of standing access with this little oversight, it would trigger an incident. Someone somewhere would say, 'Oh no, no, no, this can't happen,'" Williams says. "But the problem with NHIs is there's no one who owns the identity or who even knows it exists in some cases."

The study showed that **static, long-lived credentials or secrets are the number one most common way access is granted to NHIs and AI agents across hybrid, cloud-native, and AI-driven environments.** Organizations are more than twice as likely to use long-lived credentials (35%) as they are to use more modern just-in-time authorization (17%). And only a slim 8% use ephemeral access. On top of that, 1 in 10 organizations don't even know how they're granting access to NHIs.

Reducing standing access to NHIs and agents is a complex issue, with no single roadblock keeping organizations from modernizing. When we asked to identify the single biggest barrier, there wasn't a dominant answer that rose to the top.

Biggest barrier to reducing standing access to NHIS and agents	
Performance or reliability concerns	25%
Business expectations for speed	18%
Lack of viable alternatives	15%
Tooling or platform limitations	15%
Architectural or legacy dependencies	14%
No single biggest barrier	12%

However, all of these answers are rooted in the reality that legacy identity tools and governance weren't built for managing NHIs or agents.

59%

The study found that 59% of organizations say they lack viable alternatives to standing access for NHIs and AI agents.

Modern tooling can accelerate the path toward more sophisticated access control, but deploying modern tools won't automatically solve governance. Organizations can't build effective policies until they can observe what their NHIs are doing. Only then can they build policies and take actions grounded in operational reality rather than guesswork. In the meantime, organizations may still need to keep granting standing access to keep pace with AI innovation.

The experts say that step one should be simply identifying when and why agents are given standing access.

"I'll count it as a win if we just have an inventory that they have standing access in," Auger says.

The sentiment underscores the chicken vs. egg governance problem identified previously. **Organizations can't govern what they can't see, and most organizations lack basic visibility into their NHIs**. This is where the identity visibility is most acute in the agentic age, and until it can be closed, governance will struggle to progress.

5.0

Identity at the core of AI's biggest risks

Delinea

The risks identified by this study take on greater weight when viewed against the current threat landscape. As governance gaps emerge around machine and AI identities, attackers are increasingly targeting these access paths. AI agents are rapidly becoming part of privileged infrastructure, and unmanaged or over-privileged identities represent a high-value target. In Delinea Labs' recent report, The State of Identity, AI, and Cyber Resilience in 2026, researchers reflected:

"In 2026, the core security question is no longer 'Can we stop intrusions?' It is 'Can we continuously validate trust across humans, machines, and agents—at machine speed?'"

Threat and attack highlights:

▶ 2025 saw deepfake heists, AI ransomware, and credential leaks at scale.

▶ Attackers are themselves using AI-driven automation to infer privilege, identify high-value systems, and plan escalation paths.

▶ Ransomware tactics are shifting to target identity infrastructure, exploiting IdPs and SSO.

▶ PAM exploitation is common. Over-entitled service principals are widespread. AI weaponizes leaked identity data.

▶ Attackers target NHIs precisely because defenders cannot account for what exists, much less determine what is over-privileged.

Survey respondents are aware of these threats despite their overconfidence in agentic AI readiness. Broadly, 92% of organizations believe that AI will amplify identity-related threats over the next several years, with credential stuffing and password attacks (33%) as well as privileged account compromise (31%) leading their concerns.

92%

92% believe AI will amplify identity-related threats in the coming years

These threats and other identity supply chain attacks are raising that awareness. Delinea Labs' analysis identified a fundamental shift in how attacks unfold. Identity has become the primary execution layer. Gal Diskin, head of Delinea Labs, points to the identity supply chain attacks that closed out 2025 as evidence of this shift.

> **These attacks started from stolen credentials and propagated by stealing more credentials. The whole goal is identity. The whole method of propagation is identity."**
>
> Gal Diskin, VP, Identity Threat Product & Research, Head of Delinea Labs

Breaches increasingly originate from legitimate access—valid credentials, tokens, OAuth grants, and automation pipelines—rather than traditional exploitation of vulnerabilities.

Williams sums it up succinctly: **"Attackers don't need to break in because we are legitimately handing them access."**

Once they're in, before anyone notices, the damage is done.

"Legitimate access does not mean safe access," Diskin adds. "A lot of organizations are confident they will detect something, but detecting after the fact is very different from controlling and preventing before the fact."

This shift to runtime risk is reshaping the dynamics of identity and trusted access. Non-human identities and AI-driven systems are expanding trust relationships faster than governance frameworks can adapt—especially when agents clone workflows, share permissions, and replicate across environments.

As previously mentioned, the challenge in building a sense of urgency for governing AI identity risk is that many organizations haven't yet witnessed first-hand what failures look like. Monitoring these trends will be critical for quantifying the risks and articulating the business case to leadership. Clear visibility into identity sprawl, privilege drift, and governance gaps enables informed insight into why visibility and controls must be made before incidents force reactive action.

Auger compares it to the early days of seatbelts: "When seatbelts first came out, you didn't have evidence to make the case for putting the seatbelt on. People would think, 'Why? The car still drives.' You needed the stats that there's been this many mortalities a year to say, 'You will die in a head-on collision if you don't.' **We have to make the right case to properly convey that giving AI all this exposure is a problem.**"
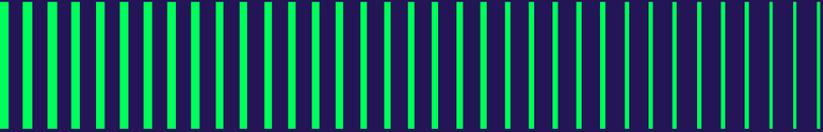
6.0

Reduce identity security friction, reduce AI risk

Delinea

Identity risk in 2026 is driven by organizations intentionally expanding trust faster than they can govern it. Attackers are using what already exists, moving faster than static controls can respond and exploiting the trust that defenders must grant to maintain operations.

The survey data tells a consistent story: Organizations are knowingly trading identity control for operational velocity. There is strong pressure to loosen privileged access requirements to enable automation. Standing access is routinely granted under operational urgency. When security introduces friction, speed frequently wins.

The report reflects a core industry truth: Organizations can't secure or govern what they cannot fully see.

Eliminating identity blind spots has become foundational to managing agentic AI risks. Organizations can't do that with human-centered or manual IAM tooling.

By automatically discovering and mapping identities and privileges across this expanded landscape, organizations can start to bring governance reality in line with confidence levels.

To further emphasize our findings, the following recommendations come from a panel of independent experts:

1. Visibility comes first

> **"If you don't have visibility into something, you can't protect it. If you don't know what's accessing your systems, you can't stop it."**
>
> Kayla Williams

It also means watching for signs of shadow AI: unusual access patterns, unexpected requests for elevated privileges, and activity that looks normal but doesn't match what you'd expect. Every other recommendation in this report depends on getting visibility right first.

---

2. Machine-speed security for machine-speed threats

Addressing identity risk in the AI era requires moving beyond human-in-the-loop controls. The scale and speed of AI agent activity make constant manual oversight fraught with friction.

> **"Machine speed is what we need. Humans are the bottleneck. There are too many agents and too many activities."**
>
> Chris Hughes

This will take investment. Without it, organizations remain stuck with legacy identity access management (IAM) tools that were never designed for autonomous, high-velocity systems.

> **"The dynamic between the executive team and the security team has to evolve into more mature and serious discussions around budgeting for modern identity management tooling."**
>
> Kayla Williams

3. Zero standing privilege is the endgame

The reliance on static, long-lived credentials for AI agents creates persistent risk. Moving to just-in-time and ephemeral access models is critical to achieving this state of ZSP.

> "Long-lived static credentials are easy to abuse. The just-in-time aspect of access control really needs to grow. We need to give agents the access, agency, and autonomy they need at the right time, with credentials that are ephemeral."
>
> Chris Hughes

Getting there requires significant identity and access management maturity. It sounds simple, but doing it right at scale is hard. For most organizations today, standing access will likely remain the realistic baseline. Which is why visibility is important to understand when and where those accounts have been granted standing access on an ongoing or temporary basis.

> "Ephemeral access is only reachable right now for, say, financial organizations that have an entire IAM team with analysts dedicated to agentic AI identities. For mere mortals in any other industry, standing access is going to be the norm. I'll count it as a win if we just have an inventory of all the identities that they have standing access in."
>
> Gerald Auger

4. Zero-trust principles are more important than ever

In some circles, zero trust is a four-letter word due to being co-opted by the security marketing buzz machine. The fact is, though, that the principles of it are more crucial than ever for accounts controlled by agents.

> **"So many CISOs hate the term zero trust, but those principles are what's going to make any of this possible at speed. Least permissive access control and microsegmentation are crucial for controlling what agents can or can't do, where they can go or not. The blast radius concept is important for limiting an agent if it either behaves in a way you didn't anticipate or even gets compromised as part of your attack surface."**
>
> Chris Hughes

One of the most central pieces of zero trust is moving to a state of ZSP, where no standing, persistent admin rights exist for agents.

5. Encourage experimentation in isolated environments

One way to balance the pressure for AI innovation with identity risk management is to provide controlled spaces for experimentation.

> **"Give people the opportunity to scratch an itch while managing the risk of exposure of your internal environment. You can give them synthetic data, public data, and allow them to tinker or innovate."**
>
> Gerald Auger

Sandboxes won't satisfy every use case. For example, Carl the accountant wants his agent connected to real systems, not a test environment. But sandboxes can channel early-stage experimentation away from production data while organizations build out governance frameworks.

Delinea

6. Evolve from least privilege to least permissive autonomy

The principle of least privilege is well established. But for AI agents, it needs to extend beyond access rights to encompass autonomy as well.

> **"Agents should only have the level of autonomy and agency they need to carry out their mission and nothing more."**
>
> Chris Hughes

This means not just limiting what systems an agent can access but also constraining what actions it can take and what decisions it can make independently.

Learn how Delinea can help
organizations better navigate
identity governance and
AI-related risks today ▶

Delinea.

Delinea is a pioneer in securing human and machine identities through intelligent, centralized authorization, empowering organizations to seamlessly govern their interactions across the modern enterprise. Leveraging AI-powered intelligence, Delinea's leading cloud-native Identity Security Platform applies context throughout the entire identity lifecycle across cloud and traditional infrastructure, data, SaaS applications, and AI. It is the only platform that enables you to discover all identities — including workforce, IT administrator, developers, and machines — assign appropriate access levels, detect irregularities, and respond to threats in real-time. With deployment in weeks, not months, 90% fewer resources to manage than the nearest competitor, and a 99.995% uptime, the Delinea Platform delivers robust security and operational efficiency without complexity. Learn more about Delinea on **Delinea.com, LinkedIn**, **X**, and **YouTube**.