Lumos

Identity Security in the Age of AI Agents

Managing the agents that are working in your organization, even when you’re not.

The attack surface moved. The toolkit didn't.

Identity is now the primary vector for breaches. Attackers don't need to break in when they can just log in, usually with credentials and entitlements that should never have existed in the first place. As AI agents spread throughout the enterprise, the number of identities acting on your behalf is growing faster than any team can manually track.

Every wave of automation has increased the identity attack surface. Workforce SaaS made access for non-employees common. Cloud infrastructure removed the perimeter. AI agents are adding a third wave: identities acting on behalf of people, often with privileged access and often without a clear owner. Each wave arrived faster than the previous one.

Each wave arrived faster than the previous one, and none of the tools designed for earlier waves had time to catch up. Most organizations still manage access like they did ten years ago: using static snapshots, performing manual reviews, and relying on tools that can answer "who has access?" but not the more critical question: "should they have it, and is it being abused right now?"

Security teams are inheriting access graphs they didn't build and can't monitor at scale, and **only find out about risky access after something goes wrong.**

What teams actually face

Silent accumulation

Overprivileged accounts and dormant access increase between reviews. By the time a quarterly review identifies a risk, that access has often existed for months, and the activity that created the risk has already occurred.

Compliance ≠ security

Existing IGA tools generate reports but lack signal about risk. They tell you who has access, but not whether they should have it or if their usage has changed in the last hour.

Siloed context

Security and identity teams work separately. The rationale behind a grant often doesn't travel with the entitlement, making it difficult for the team managing an incident to reconstruct the context under pressure.

Event-rich, identity-poor.

Detection tools flag actions but lack entitlement history to determine if an action is expected or unusual. Alerts pile up, and the signal-to-noise ratio worsens as the environment grows.

Why the existing toolkit can’t close the gap

Most organizations are managing identity security with tools designed to tackle different problems. Each category does its part, but none are made to govern the entire identity attack surface, both human and non-human, in real time.

Each tool has features that seem related: the IGA platform reports on entitlements, the SIEM correlates events, the PAM tool monitors privileged sessions, and the ITDR vendor oversees authentication. However, none can determine whether a specific human, service, or agent should have a certain entitlement at any given time. Therefore, they can’t answer that question without human intervention to piece together the story.

Tool category	Where it stops short
Traditional IGA	Focuses on compliance rather than security. Built for audit cycles, not threat response. Reports appear thorough, but the risks often remain hidden.
PAM	Strong for privileged accounts but overlooks the many SaaS and app entitlements where most of the overprivilege actually exists.
SIEM / SOAR	Rich in events but lacking in identity context. It cannot assess an action against entitlement history, so anomalies are left uncorrelated with the access that enabled them.
Manual review programs	Provide quarterly snapshots in a constantly changing environment. Changes between cycles remain unseen until the next review.

The gap is not about any single tool. It lies in the lack of a continuously updated, security-grade view of the entire identity environment and the model needed to act on it. Closing the gap requires linking entitlement data, usage behavior, and risk signals in one system. This way, detecting overprivilege, evaluating it in context, and fixing it can all happen in the same workflow, using the same data, in a timely manner.

A new category for a new attack surface

In its 2025 Hype Cycle for Digital Identity, Gartner introduced a category that addresses what the current toolkit lacks: Identity Visibility and Intelligence Platforms: IVIP. While IGA, PAM, and SIEM each cover parts of the identity environment, an IVIP provides rapid integration and unified visibility across an organization's complete IAM infrastructure–human, non-human, and AI-agent identities–all in one place.

This category offers a useful label for what every identity team has been piecing together. It also acknowledges, as stated by an analyst, that the piecemeal approach to identity security is no longer sufficient. Visibility and intelligence are fundamental components. Almost every modern identity program needs these traits, and the number of tools offering these features is growing.

But even in a category as recently created as IVIP, the speed of AI adoption has already begun to outpace new platforms built on IVIP principles. This is because many of them aren't built for the speed at which AI agents proliferate, and even those that are still treat AI NHIs as a separate category to manage.

The Closed Loop: See, Understand, Act

In the modern AI agent era, stopping at visibility and intelligence falls into much the same trap that a SIEM falls into. You end up with a high-resolution picture of your risk and the same backlog you started with; better informed, no more secure. The solution is a third step: Action. The closed loop completes the picture in three steps, while most platforms cover only the first two.

See

A continuously updated map of every identity, entitlement, and behavior across human, non-human, and agent identities, all in one place.

Understand

AI-ranked findings with straightforward explanations that describe why a set of factors is risky, not just that they exist.

Act

Identity Security Agents that detect, reason, and remediate, at the level of human oversight you choose, and turn fixes into enduring governance policies.

That last step is where many identity tools fall short, as they don't manage the lifecycle, access reviews, or provisioning policies needed for sustainable solutions. It's the difference between revoking the same over-privileged admin every quarter and creating a policy that stops the grant altogether. Identifying a risk once and fixing it permanently; that is what closing the loop really means.

A new operating model for identity security

Modern identity security isn't just another dashboard added to outdated data. It's a continuously updated map of every identity, entitlement, and behavior, combined with AI agents that actively search for risks and remediation processes that occur on the same platform.

The shift is from viewing access as a configuration to seeing access as a continuously evaluated aspect of the identity. Configurations are reviewed quarterly, while characteristics are monitored every minute. This change significantly speeds up the detection of risky access, boosts confidence in remediation, and lowers the human effort needed for each decision.

Full attack surface visibility

Human and non-human identities—service accounts, automations, AI agents—across every connected app, displayed in one access graph.

Proactive risk hunting

AI agents identify anomalies, flag harmful combinations, and uncover dormant privileged access without waiting for a review.

From finding to fixed

Remediation happens where the finding is identified. Detect, assess, and revoke risky access on the same platform—no separate ticket needed.

Identity intelligence everywhere

Risk signals and access context flow into the tools security teams already use, making identity part of every investigation, rather than a separate one.

What this changes for the team

Security gains a constantly updated view of the full identity attack surface. IAM benefits from automation that helps manage the noise. Compliance receives documentation as a byproduct of ongoing work. The organization gains the ability to detect and respond to identity risks in hours rather than quarters.

It also changes what "good" looks like for an identity program. Rather than measuring success by completion rates of reviews, teams can measure it by the level to which access aligns with polic,y and the time taken to address a risky entitlement. Those metrics are easy for a board to understand and hard for an attacker to overcome.

The new identity type: agents

Every wave of automation has expanded the identity attack surface, and the current wave is advancing faster than the controls. Non-human identities like service accounts, API keys, OAuth grants, and workload credentials already outnumber human identities in most enterprises by 10:1 or more. AI agents are introducing a third wave: identities that act on behalf of people, often with privileged access and frequently without an identified owner.

AI agents are a third identity category

Until recently, identity governance primarily dealt with two categories: humans and machines. Humans possess judgment; machines have computational power. AI agents are different: they combine reasoning and judgment with the scale of machines. A human with read-write access to a CRM might update a few records daily. An agent with the same access can export an entire database in seconds.

This changes what governance needs to achieve. Authentication reveals "who" the agent is. Authorization specifies what the agent can do. Neither answers the tougher questions: why does it have that access, who approved it, is it appropriate, and what happens when it's not? The answers to these governance questions are what is lacking in many agent deployments today.

Agent access also rarely takes one form. It may be granted as service accounts in cloud IAM, OAuth grants in SaaS, API keys in developer tools, certificates and workload identities in infrastructure, or increasingly, MCP connections that inherit a human user's full permissions. Each access method has different control surface, and none were designed with the presence of an AI agent in mind.

Three fundamentals that close the gap

Across hundreds of customer environments, the same three key principles keep separating identity programs that govern agent access from those that just inventory it.

Every connection has an owner

An agent connection without an owner is one that lacks governance. You cannot review access that nobody owns, enforce a lifecycle for it, or trace intent back to a decision that lacks a clear source.

Intent travels with the credential

When access is granted, the stated purpose is recorded along with it. Without this link, monitoring has no baseline to compare against.

Composite access is reviewable

Each access scope may seem reasonable on its own. However, the complete set granted to one agent—such as Slack history, GitHub repo access, and Sentry issues—needs governance.

Identity security agents: The action layer

This is where Identity Security Agents become important. They serve as the action layer of an IVIP, designed for specific tasks and capable of monitoring, detecting, and responding to risks without waiting for human intervention. If Albus is the analyst on your team, handling investigations, pattern recognition, and recommendations, then Identity Security Agents are the operators acting on that information. Albus provides insight into the data; the agents take action based on it.

Two of the initial agents Lumos has introduced address urgent gaps. The NHI Security Analysis agent scans Microsoft Entra ID to identify complex risk patterns—combinations like privileged access, multi-tenant exposure, long-lived credentials, and no assigned owner—that a single signal won't detect and that a manual analyst would take weeks to uncover. The Terminated User Account agent checks HRIS termination data against active access in connected systems, revealing post-departure access that was missed during offboarding. Both agents provide findings with complete evidence and suggested actions; they also learn from your team's feedback over time.

The model relies on progressive trust. Agents start with full oversight, where every recommendation surfaces for review before anything executes. As your team builds confidence, more is delegated. Initial calibration happens in a single session, not weeks: install, run a test scan, dismiss noise with feedback, re-run, repeat until the agent’s output matches your team’s expectations. Every dismissal becomes part of the baseline, and the agent gets sharper with each pass.

And because the agents live on the Lumos platform, the loop does not stop at remediation. A finding the agent surfaces and the team confirms can be translated into a durable governance rule — enforced through the same access reviews, lifecycle workflows, and provisioning policies that already govern human access. The same fix does not have to be made twice. That is what it means for IVIP, Security Agents, and IGA to live in one place.

A starting checklist for security leaders

- Can you quickly generate a current inventory of human and non-human identities, including AI agents and their connections? How long does this take?
- Can you show, on demand, where overprivileged access lives and how it changed over the last quarter?
- Is risky access detected continuously, not just at review time?
- Does every agent and non-human identity in your environment has a named owner — or a triageable workflow to assign one?
- How quickly can findings be remediated inside the same platform that surfaced them, and then turned into governance policy that prevents recurrence?
- Are identity risk signals flowing into the tools your SOC and IR teams already use?

See it in action

Lumos brings IVIP, Identity Security Agents, and full IGA together in one platform — so security teams can see what they couldn’t before, understand which combinations create real risk, and act before an attacker does. Request a personalized walkthrough at lumos.com.

Book a Demo