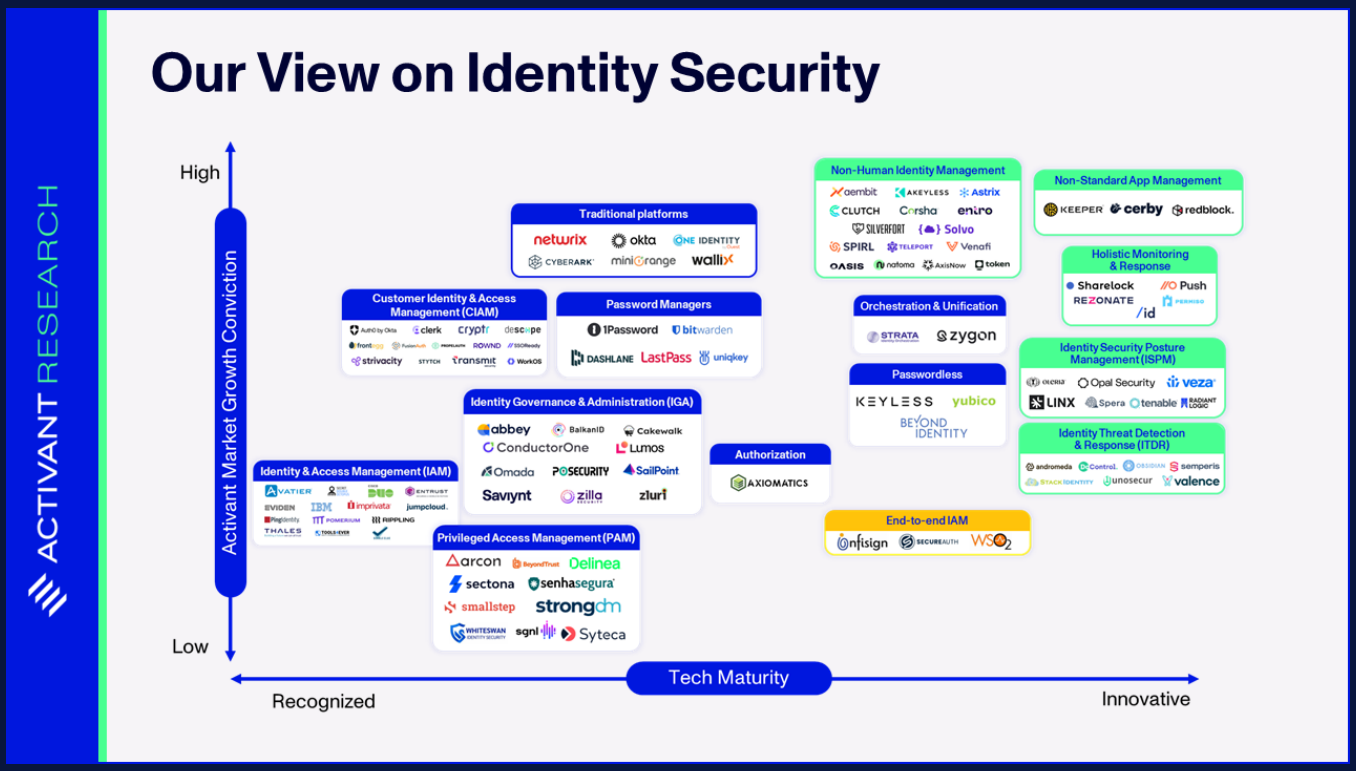
ACTIVANT RESEARCH

Authenticating AI Agents

The Next Big Identity Crisis in Cybersecurity



Gregory Rawbone-Viljoen, Jonathan Vickery

Q1 2025

Knowledge work, such a critical piece of the modern economy, is increasingly executed via software applications. Humans access these applications and their underlying data using credentials, like a username/password pair. Identity Security, a $76bn market, has been established to secure these human identities. However, increasingly large swaths of today's work are not done by humans, but by software.

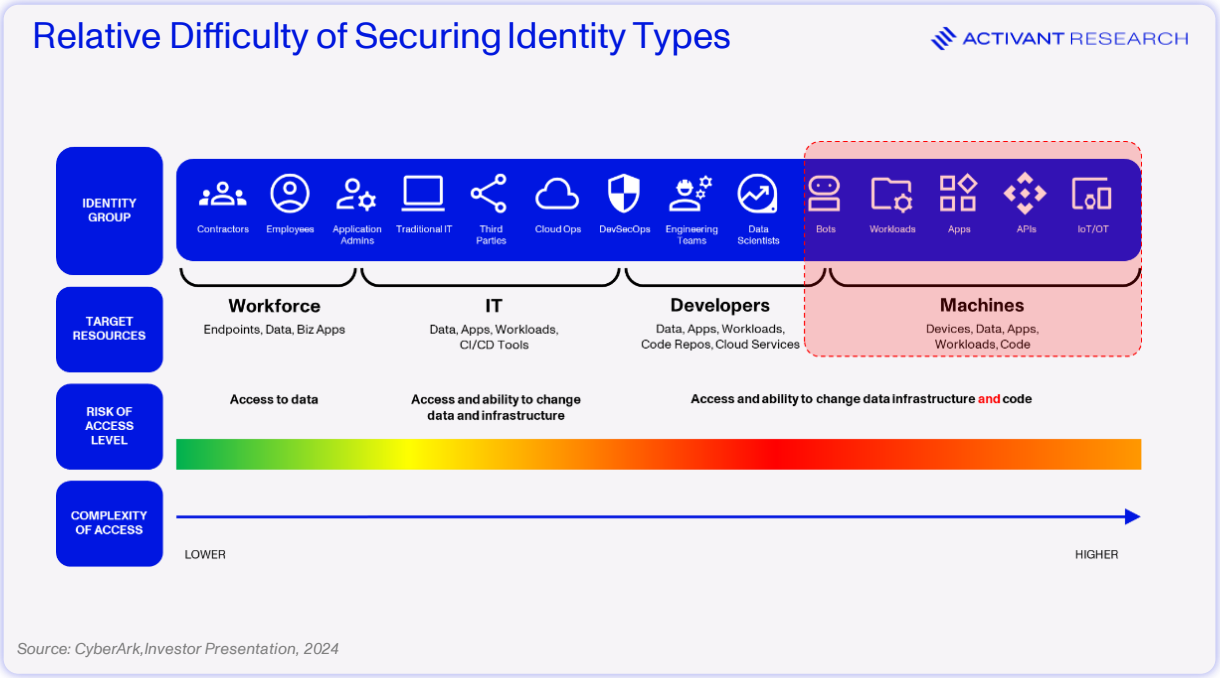
Cloud workloads automatically provision new infrastructure and test and deploy code; software applications automate business processes such as payroll calculations; and bots execute trades in financial markets. Just like humans, these applications can't have uncontrolled access to enterprise resources. We need non-human identities (NHIs) like API keys, X.509 certificates, and OAuth tokens to control which machine entities access specific data, and under which circumstances. Think of it as identity and access management (IAM) for machines.

However, as covered in our recent article, non-human identities are difficult to secure and quickly becoming a key attack vector. Fifty-five percent of organizations have experienced a breach or incident via their software supply chain that exploited vulnerabilities in NHIs.[1] Driven by cloud adoption, business processes automation and a shift to microservices architectures, NHIs are growing at a rapid rate of ~2.5x every year, outnumbering human identities by a factor of 45x.[2] Yet, the amount spent on NHI security is only 5%–10% of the $76bn spent on human identity security.[3] If that sounds concerning, consider this: enterprises are racing to adopt AI agents, and soon every SaaS app will launch its own AI agent feature. **We are about to witness a massive acceleration in the proliferation of NHIs.**

In this article, we examine the impact of AI agents on NHI security, explore emerging authentication methods, and offer our recommendations for the future.

Identity's Blind Spot: Non-Human Identities

IAM systems ensure that only the right people – or machines – have access to the right enterprise resources, for the right reasons. However, not all identities are alike. The need to access data and make changes to core infrastructure makes machines, or non-humans, a complex and risky area of IAM.



Source: CyberArk,Investor Presentation, 2024

NHIs are digital representations of these non-human entities and includeapplications, services, APIs, virtual machines, containers, bots, and IoT devices. NHIs enable secure machine-to-machine communication. In this context, the identity refers to the non-human entity registered in an IAM system, while the associated digital credentials (often called secrets) verify that identity. These digital credentials can take the form of certificates, tokens, or keys.

API keys are among the most recognizable NHIs, but their static nature carries security risks. OAuth tokens are likely the most prevalent, and users will recognize their fine-grained authorization when granting services access to view Gmail account details for single-sign-on purposes. However, OAuth tokens are also susceptible to theft and hijacking and they can be challenging to manage in distributed systems – since their expiration times are fixed at creation, revocation across widespread systems becomes difficult. Certificates provide the most robust security but require costly and complex management infrastructure. As a result of these trade-offs, the industry has standardized on using OAuth tokens for most use cases, and the security risks that this presents are critical for understanding the deeper issues posed by NHIs for identity security.

Key Non-Human Authentication Methods

ACTIVANT RESEARCH

Authentication Type	Common Examples	Benefits	Weaknesses
Certificates (X.509)	• Secure web communications (HTTPS/TLS) • Email (S/MIME) • Code/software signing • IoT device authentication	• **Strong security using asymmetric cryptography** • Supports bi-directional authentication • Allows for centralized management • Provides integrated encryption	• **Complex set up and management** • Can be slow to revoke • Higher computational requirements • Can face interoperability issues
Tokens (OAuth, JSON Web Token)	• API/microservices communication • Cloud service integrations • Single sign-on (SSO) • Third-party integrations	• Stateless (no server-side storage) design is simple, decentralized & scalable • OAuth standard is interoperable • **Fine-grained access control**	• **Susceptible to token theft/interception** • Difficult to revoke before expiration • Requires token lifecycle management
Keys (API Keys, SSH Keys)	• Internal service APIs • Cloud integrations • Automated processes and scripts	• **Simple and quick to implement** • Low computational overhead • Ubiquitous support	• **Static nature poses security risk** • Vulnerable if intercepted without secure transport • Lack of fine-grained access control • Difficult to revoke

Consider the following example: a user approves access for a project management app to read their contacts from their core productivity suite (likely Microsoft 365). The flow is relatively simple and ubiquitous across many apps, highlighting the innovation that the OAuth standard has brought to the identity market. However, note that in this example (as in many real-world cases) the token is valid until manually revoked.

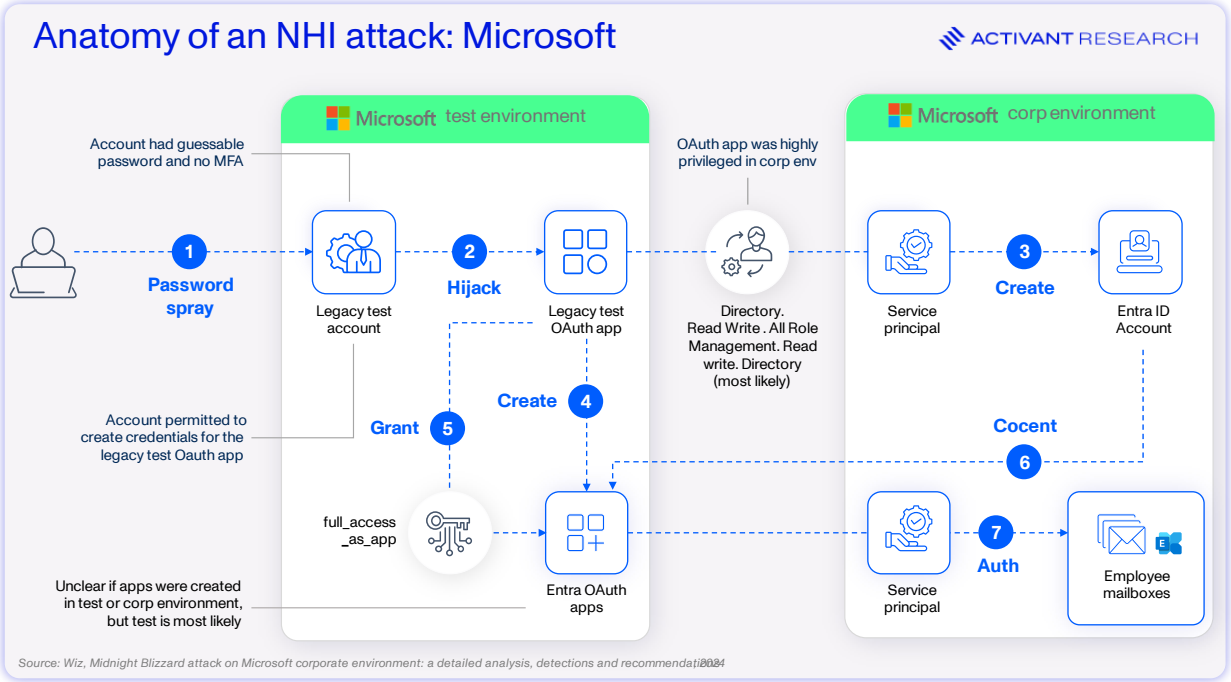
OAuth Token Access Flow Example

ACTIVANT RESEARCH

Source: AppOmni, OAuth Token: What It Is, How It Works, and Its Vulnerabilities, 2024

Because tokens and other authentication mechanisms require manual revocation, proper NHI lifecycle management is essential, yet is rarely implemented. As noted in our previous article, the lack of proper lifecycle management is just one of many pain points faced by those trying to secure NHIs. Additional issues include poor secrets management, over-permissioned access, secret sprawl, and the inability of NHIs to adhere to human-centric security measures. Unsurprisingly, hackers are increasingly exploiting these vulnerabilities – nearly one in five organizations has experienced an NHI-related security incident.[4] Some of the world's most renowned software enterprises, including Dropbox, Cloudflare and Microsoft, have not been immune to such incidents.

In the case of Microsoft's attack by an alleged Russian state-sponsored actor known as Midnight Blizzard, attackers were able to leverage an NHI – a legacy OAuth test application – to gain access to Microsoft's IT environment. Because the account possessed excessive privileges, the attackers were able to create new applications and user roles, move laterally within Microsoft's IT environment, and eventually access Microsoft employee emails. While the details are technical, the key takeaway is clear: **NHIs are over-permissioned and lack MFA, visibility and oversight, as well as proper lifecycle management for timely deprovisioning.**



Adequately securing NHIs is one of the biggest, if not the biggest, challenge in identity security today – only 25% of organizations believe that they can prevent an NHI-based attack. **We've said before that identity is the fundamental flaw in cybersecurity, and NHIs are the Achilles' heel in this space**. The most critical consideration of all is that we find ourselves in this situation before the inclusion of AI agents in our IT landscapes.

AI Agents: A New Threat

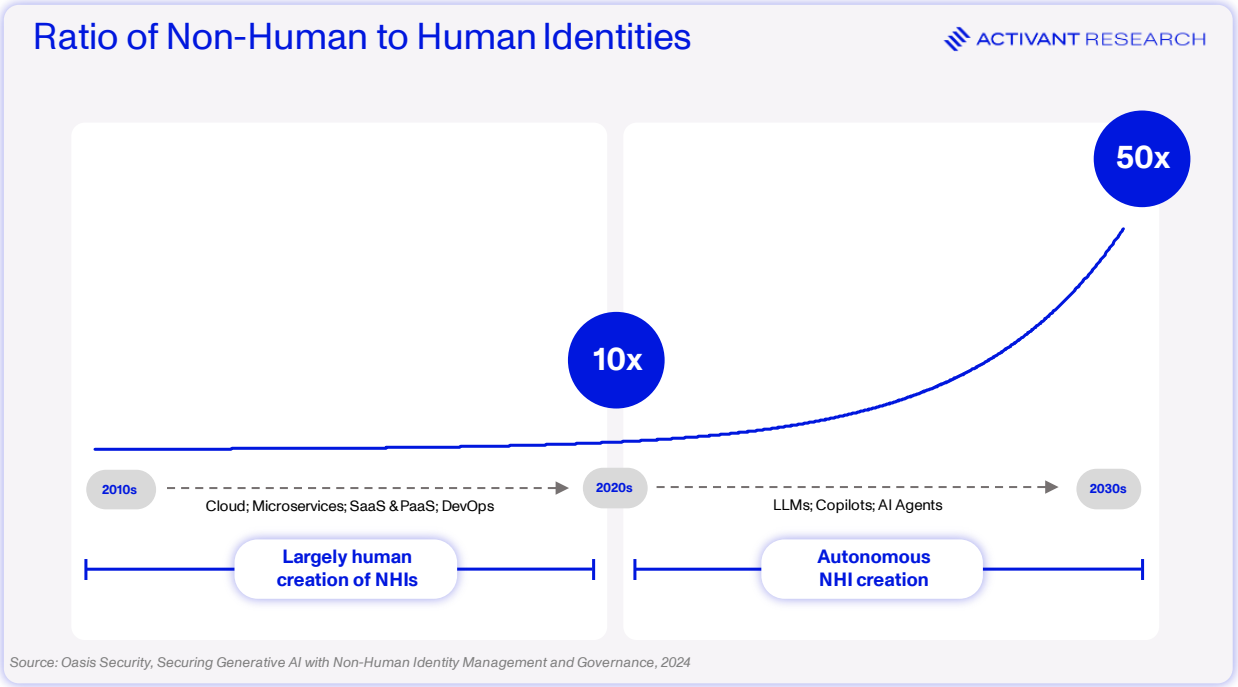
Salesforce rebranded their AI Copilot as Agentforce, Microsoft CEO Satya Nadella declared that agents are the future of software, and McKinsey has already published thought piece on why AI agents are the next frontier of Generative AI. AI agents are happening right now, and adoption is moving incredibly fast. Expectations are that 82% of organizations will implement AI agents in production within three years, with AI agents expected to make 15% of daily business decisions.[5,6]

Of course, AI agents are a major security consideration: 84% of companies see cybersecurity as the major roadblock to adopting AI.[7] To bring AI agents into the enterprise in a way that does not expose major cybersecurity risks, they must comply with a modern zero trust approach, which means starting with authentication. For that reason, AI agents will soon become the major focal point of NHI security, introducing some fundamentally new risks.

To complete their work, these AI agents will require access to a wide range of applications and data. For example, a sales outreach automation agent might require access to CRM systems like Salesforce and may need to query transaction history and engagement metrics in a database like Snowflake. The agent would need to integrate with email marketing platforms like MailChimp, access a salesperson's Outlook calendar for scheduling, and integrate with an LLM provider like OpenAI to generate emails. This simple example could result in 4 OAuth tokens (Salesforce, Snowflake, MailChimp & Outlook) and one API key (OpenAI). Finally, imagine a scenario where AI agents can spawn additional AI agents, exponentially increasing the number of credentials and potential attack vectors.

> "A single misconfigured AI Agent can generate dozens of privileged NHIs per day, each with unknown risks, significantly expanding the identity attack surface"
>
> **Oasis Security**

Furthermore, the speed at which companies are racing to adopt AI agents, either DIY or via SaaS offerings, is going to create security vulnerabilities in and of itself. **AI agents aren't just a new technology – they're a new attack vector.** A fast-paced implementation of an experimental technology with wide-ranging enterprise access privileges may just be a CISO's worst nightmare.



But the concerns surrounding AI agents go well beyond the fact that there will soon be a great many of them and there are fundamental reasons to be more concerned about AI agents than other NHIs:

1. **Susceptibility to social engineering:** Previously, only humans were vulnerable to social engineering and breaches typically exploited software vulnerabilities. In contrast, AI agents introduce new risks: they are susceptible to both software exploits and social engineering. For instance, researchers managed to 'jailbreak' ChatGPT with a 90% success rate and, in another experiment, users convinced an AI system to transfer funds against its explicit programming, netting one researcher $50,000 in just five minutes.[8,9,10]
2. **Complex, unpredictable chains of events:** AI agents work through iterative loops that involve multiple system interactions, creating a massive surface area for attack where every single system interaction is a potential vector for data exfiltration or code injection. More importantly, the unpredictability of these interactions might render traditional analytics and anomaly methods of detection ineffective.
3. **Dynamic use cases:** Fundamentally, AI agents operate stochastically and their actions may show large variance. For this reason, granting AI agents static access privileges is inappropriate. An email agent may need rights to send emails when it is responding to basic items like scheduling requests, but those send privileges should be revoked when the agent is handling sensitive information.

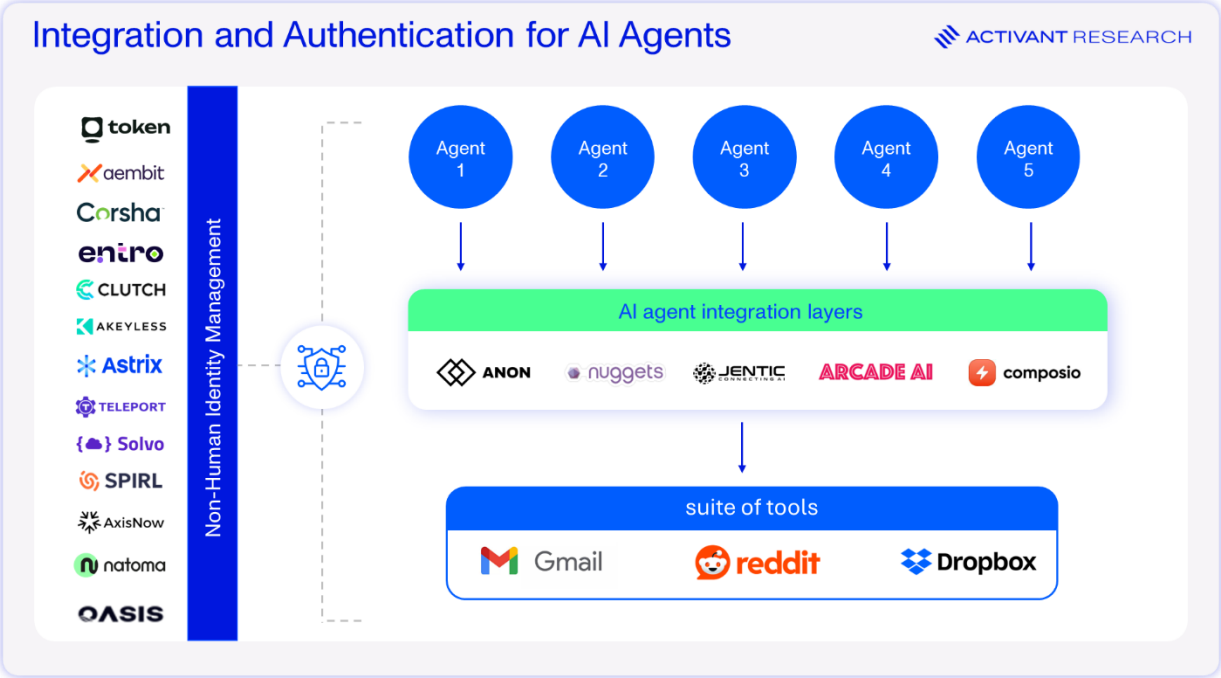
As a result, 25% of breaches are expected to be traced back to AI agent abuse in the next three years.[11] **When considering that most companies haven't even set up MFA for their human identities, it's fair to say that these agent-driven breaches are inevitable**. Companies need to be

proactive in adopting best-in-class security tooling, purpose-built for AI agents to prevent becoming part of the statistics.

How to Authenticate the Next Generation of AI

When it comes to authenticating and authorizing AI agents in enterprise environments, the fundamentally new issues that we've outlined above would lead some to suggest that we need brand new identity security tooling. Others have dismissed these concerns, believing that AI agents are just another machine entity that can be handled with our existing tooling, like NHI management.

Non-Human Identity Management



Companies like Aembit, Entro, Astrix, Oasis, Token Security and Teleport are innovating quickly to solve the core issues related to NHIs that we outlined earlier. For AI agents, these tools offer the capability to quickly scan for NHIs and discover all AI agents, taking steps to manage the problem of agent proliferation. If a SaaS app releases an AI a and an employee provisions it with excessive access privileges, users of NHI management software will not only have visibility of the issue but will be appropriately alerted by risk prioritization scores. These companies also offer NHI lifecycle management and offboarding tools – so AI agents' privileges can be revoked when they are no longer needed. Finally, if a third party system suffers a breach related to an AI agent, these tools offer the capability to flag the issue and automatically remediate by rotating any secrets that were exposed to the third party.

While AI agents pose new and concerning security risks, we believe that existing NHI management players are well positioned to respond to the risks. Ultimately, the security concerns associated with AI agents may further place non-human identity security into focus and accelerate adoption, making these companies massive beneficiaries of the adoption of AI agents.

> AI agents are being adopted at an unprecedented pace – without the necessary guardrails or security. An automated process that isn't documented, controlled, or deterministic is a severe risk. Identity will be a key solution. We must ensure that AI workloads are constantly authenticated. AI agents are going to accelerate identity security like the shift to mobile accelerated Uber.
>
> **Ido Shlomo, Co-founder & CTO, Token Security**

But, as we've flagged fundamentally new issues associated with AI agents, will it be enough? The stochastic nature of AI agents may break traditional risk scoring and anomaly detection, and rules-based policy workflows used for lifecycle management will need to be updated to be dynamic. Passwordless authentication is creating a new authentication mechanism for humans, rather than trying to use layers of identity security posture management (ISPM) and identity threat detection and response (ITDR) to deal with the issues associated with username and password pairings. In the same way, we may need a new authentication paradigm for AI agents that confronts their underlying security risks head-on, rather than bolting on discovery and threat alerting afterwards.

Integration & Authentication Tools for AI Agents

The key security risk for AI agents is the number of data and systems that they will have access to. Tools like Composio, Anon, Jentic and Arcade are building the integration layer to control those data and system connections, which makes them a focal point for AI agent authentication. Anon is building integration tools that authenticate AI agents, allowing them to take actions on behalf of users on the internet. For example, a user would use Anon's services to use an AI Agent to scan Amazon.com for pricing data on a product category they're researching. This approach obviates the need to build complex logic for web scraping and data normalization, all while leveraging an authenticated user identity via OAuth or SSO. Anon links a user's account and passes a human identity on to the AI tool in question. Similarly, Composio and Arcade provide a suite of integrations that allows AI agents to authenticate and connect, via API, to tools like Salesforce or Github. Both tools can leverage OAuth, SSO or JWTs and hand the user account to an AI agent to take actions on the user's behalf.

These companies are building the early infrastructure for AI agent tooling that conforms with Zero Trust principles, but we see multiple issues with the current state of the technology. In all cases, these agents are leveraging a static human identity with no lifecycle management. Over the longer term, this is going to expose AI agent implementations to many of the shortfalls that current NHI security programs are struggling with. Applying human identities to AI agents will leave many

overprivileged and a lack of lifecycle management will mean that stale access privileges are never revoked, becoming an attack vector. So, what will the future of best-in-class AI agent authentication look like?

Where we see AI Agent Authentication Going

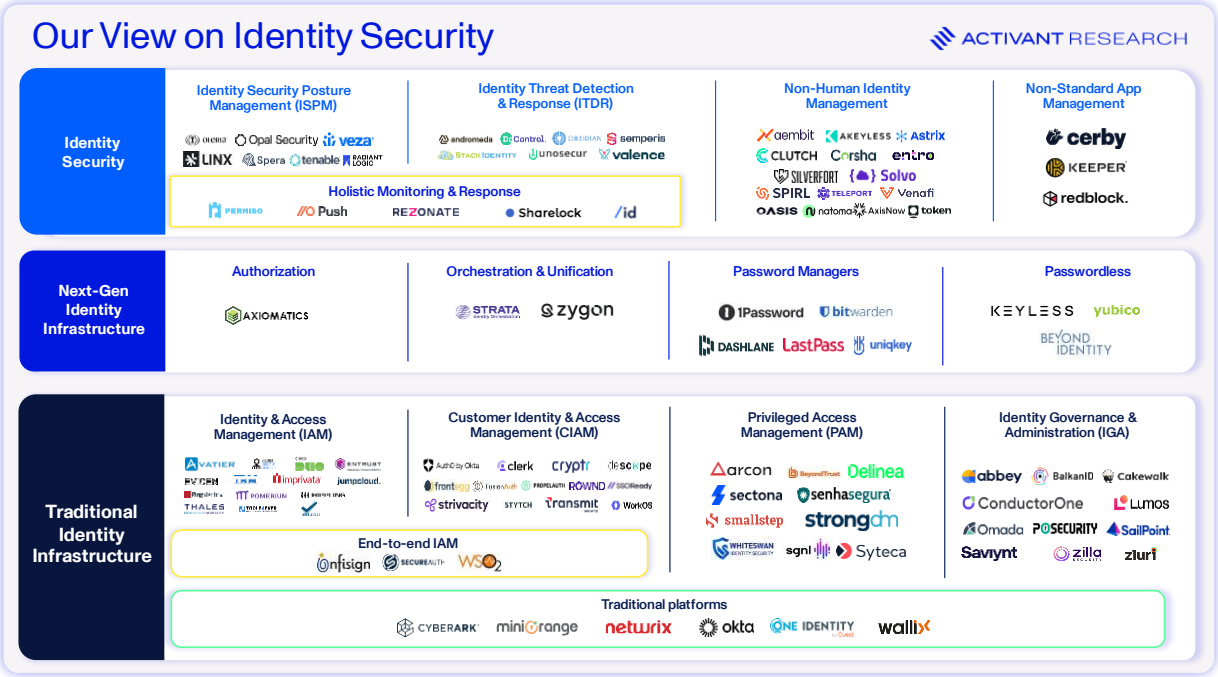
Future solutions need to fuse strict security with the dynamic nature of AI agents. We see four key areas that will define the future of AI agent authentication:

1. **Distinct Identities** will enable improved tracking, auditing and granular authorization policies. AI agents that leverage the identity of their human counterparts will soon fall away.
2. **Context-aware, just-in-time (JIT) and short-lived access privileges** should be applied. AI agents should not have standing access to sensitive resources. IAM systems should be able analyze an AI agent's access request and only grant privileges when contextually appropriate. Aembit's secretless identity tool for workloads could be extended to AI agents to serve this exact feature requirement, positioning the company well for the agentic era.
3. **Modern monitoring and alerting** systems need to constantly monitor the use of AI agents in an enterprise IT environment. Traditional anomaly detection might be unsuccessful and the rate at which these logs will be generated will make it impossible for humans to parse. The solution might be using ***more*** AI agents to constantly watch the actions of existing AI agents, and flag behavior that they consider to be a security concern.
4. **Limit sensitive data exposure risk** with data security best practices, such asencryption, secure collaboration tooling, and masking tools like Skyflow's Gatekeeper.

Once the market has moved to apply what we see as these mission-critical features, AI agents may continue to drive significant change in the broader identity security market.

Identity's next Billion-Dollar Question(s)

Looking at the existing market for identity security, we see AI agents having the potential to drive a few fundamental shifts.



AI agents continue to blur the lines between human and non-human identities, an area that is already murky. Managing this issue is going to place greater pressure on already fragmented identity tech stacks (94% of organizations rely on ten or more identity security vendors[12]). The market will be pushed to adopt modern solutions that can unify both human and non-human identity management and provide tooling across the identity infrastructure and security spectrum. The space is deeply in need of a true platform offering and we see the market consolidating on the vendor who can provide this. As we mentioned in our last piece, Okta and Microsoft are best positioned to step into this role, but the move towards agentic software could shift the competitive center of gravity.

Identity is a $76bn TAM, and massive businesses have been built by providing the critical infrastructure that sits at the center of the identity stack. Microsoft's Azure AD became the on-premises leader by providing a leading IdP, and Okta built a $16bn business by transitioning IdP to the cloud. AI, and agentic computing, represent the next major platform shift in enterprise computing and **the leader of the shift to AI-native authentication may in time become the central piece of modern identity**. The next Okta might start off in what is today the small area of agent authentication.

Finally, while we've focused on AI agents as a risk vector, there is a huge opportunity to use AI agents for good, not just in IAM and but across security. As one example, the idea of a guardian agent has been proposed, meaning that AI agents will be deployed to autonomously track, oversee and contain the results of AI agent actions. Twine is rolling out a suite of AI agents to assist in automating security tasks, and their first, Alex, is focused on IAM. Alex can perform tasks such as enforcing MFA and cleaning up stale accounts. Meanwhile, companies like Backline, Dropzone AI

and 7AI are building AI agents for other repetitive security tasks including analyzing and prioritizing alerts, investigating threats and remediating simple issues. **AI agents might be security's biggest risk, but also its biggest opportunity.**

Capitalizing on a Critical Moment in Identity

Identity Security stands at a critical moment as AI agents pose the dual risk of rapidly expanding the surface area for cyber attacks and presenting new risks that traditional measures may be unable to address. The market must quickly implement dynamic, context-aware authentication, based on distinct machine identities. We see this as an incredible opportunity for the software vendors who can adapt their solutions for the agentic world, leverage agentic security as a new center of gravity in identity, and use AI agents for good.

If you are building in the space or have an alternative view on the sector, then please get in touch.

We would love to hear from you.

Endnotes

[1] Forrester, Predictions 2022: Cybersecurity, Risk, And Privacy, 2021

[2] CyberArk, Key Considerations for Securing Different Non-Human Identities, 2022

[3] Activant Ecosystem Interviews

[4] Astrix & Cloud Security Alliance, The State of Non-Human Identity Security, 2024

[5] Gartner, Intelligent Agents in AI Really Can Work Alone. Here's How, 2024

[6] Capgemini, Harnessing the value of generative AI, 2024

[7] IBM, What Generative AI means for your data security strategy in 2024, 2024

[8] Deng. Z. et al, AI Agents Under Threat: A Survey of Key Security Challenges and Future Pathways, 2024

[9] Jailbreak refers to scenarios where users deliberately attempt to deceive or manipulate AI agents to bypass their built-in security, ethical, or operational guidelines, resulting in the generation of harmful responses

[10] Medium, How Someone Won $50,000 in 5 Minutes by Making AI Hallucinate: A Groundbreaking Security Find, 2024

[11] Gartner, Gartner Unveils Top Predictions for IT Organizations and Users in 2025 and Beyond, 2024

[12] CyberArk, Identity Security Threat Landscape, 2024