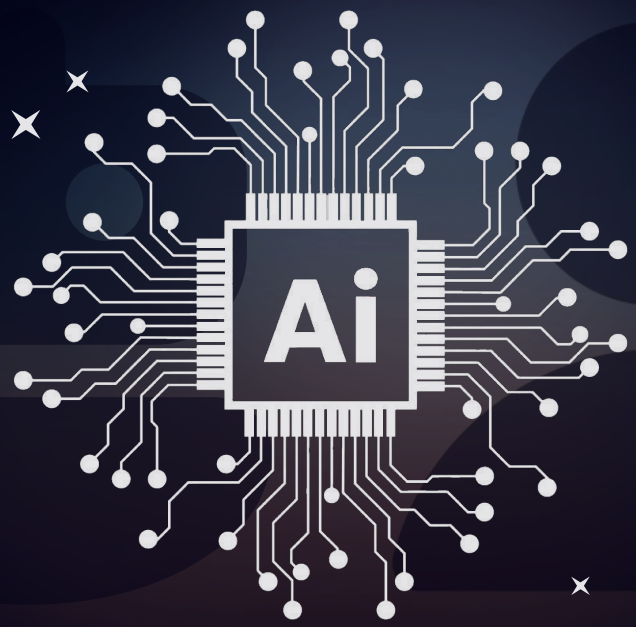
OASIS

AI Agents:
Human or Non-Human?



Overview

Introduction

During CES 2025, Jensen Huang (CEO of NVIDIA) stated in his keynote:

"... In the future these AI agents are essentially digital workforce that are working alongside your employees doing things for you on your behalf, and so the way that you would bring these special agents into your company is to onboard them just like you onboard an employee".

This vision raises a fundamental question for Identity Security: How do AI agents fit into the IT environment? Should they be managed like human employees - with centralized oversight, defined roles, and governance structures, meaning they have assigned job codes in an HR database? Or should they be treated like workloads, relying on decentralized non-human identities (NHIs) for authentication and operations?

At first glance, AI agents seem like digital employees - they assist with IT support, cloud optimization, automate customer service, content creation, and even decision-making. However, they differ from human employees in several critical ways:

- **AI agents don't have intent:** They execute tasks based purely on logic and objectives, without human reasoning, even though advanced models are becoming increasingly capable of human-like decision-making.
- **They don't use usernames and passwords:** Instead of traditional authentication methods like passwords, with compensating controls like MFA or SSO, AI agents rely on API keys, managed identities, service principals, and other machine-to-machine authentication methods.
- **They lack contextual awareness:** Humans naturally apply judgment and ethical reasoning when making decisions. AI agents, however, strictly follow instructions, meaning they may misinterpret incomplete, ambiguous, or misleading context, leading to hallucinated outputs, unintended actions, or even security incidents - sometimes with significant consequences.

These traits clearly demonstrate that AI Agents require specialized governance to prevent risks like privilege escalation, credential sprawl, and unauthorized actions.

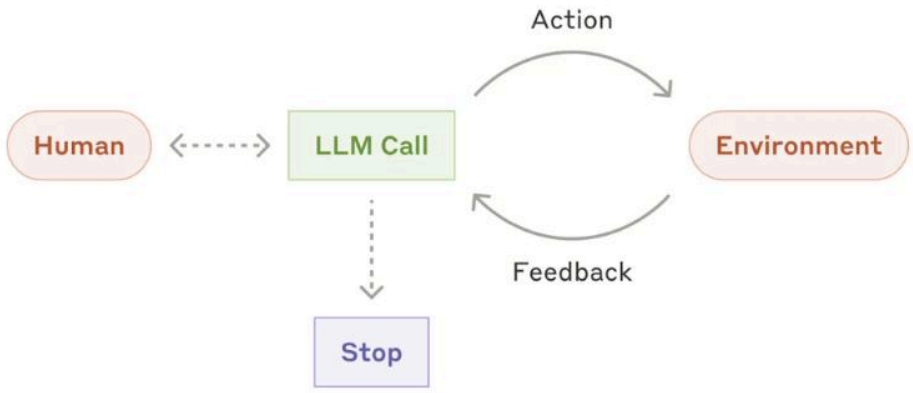
What is an AI Agent?

Many people assume an AI agent is just a piece of software that performs some operations and, at some point, makes an API call to an LLM (Large Language Model). But this is not exactly right.

AI Workflows vs. AI Agents

According to Anthropic, the key distinction is:

- **Workflows** are structured systems where LLMs and tools follow predefined code paths. These systems have clear steps, making API calls to an LLM at specific points.
- **AI Agents** are dynamic systems where LLMs direct their own processes and tool usage, deciding in real-time how to accomplish a task. In other words, AI Agents maintain control over how they accomplish tasks.



AI Agent pattern. Source: Anthropic

AI Agents start with a user command, discussion or objective, then plan and operate independently, only seeking human input when needed. They rely on real-time environmental feedback to track progress and may pause for human review at key points. Tasks end upon completion or predefined stopping conditions.

In short: A workflow is predictable; an AI agent adapts, iterates, and makes independent decisions based on its environment.

If you want to dive deeper, the foundation of AI agents lies in LLMs enhanced with retrieval, tools, and memory. Learn more in our blog: Securing Generative AI with Non-Human Identity Management and Governance.

Why AI Agents are not human employees

Despite their human-like analogies, AI agents cannot be managed like human employees. Key differences include:

- **Authentication is different:** As covered earlier, AI agents rely on API keys, tokens, or managed identities, which, if not properly managed, can lead to risks such as credential sprawl, hardcoded secrets, and privilege creep.
- **No clear ownership:** Unlike human employees, AI agents don't have a designated owner responsible for their actions - the agent often acts on behalf of an application or business vertical. This makes accountability and oversight more challenging.
- **Lack of structured access control:** There is no standardized process to enforce least privilege, meaning AI agents may accumulate excessive permissions over time.
- **No defined offboarding process:** AI agents don't follow a structured lifecycle, and there is often no formal process to revoke their access when they are no longer needed.

This autonomy is what makes AI agents so useful, but also introduces major security challenges - especially when they can operate without centralized governance.

Let's walk through an example, imagine the following scenario: AI Agent for Cloud Cost Optimization in Azure.

AI Agent for cloud cost optimization in Azure

A company deploys an AI agent in Azure AI Foundry to help optimize cloud costs by analyzing underutilized virtual machines (VMs) and automating resource scaling. Initially, the AI agent is granted read access to Azure billing data and monitoring logs to suggest cost-saving opportunities.

Step 1: Developers expand the AI Agent's capabilities

After seeing promising results, developers decide to let the AI agent take action instead of just making recommendations, granting write access to:

- Stop or scale down underutilized VMs to reduce costs.
- Generate cost reports and update internal dashboards.

To make this happen, the AI agent is integrated with Azure Automation and Azure Functions to execute cost-saving workflows.

However, due to misconfiguration, the AI agent also receives broader write access than necessary - including the ability to modify identity and access management (IAM) policies.

Step 2: The AI Agent requests more access

Once the AI agent starts executing cost-saving measures, it runs into permissions restrictions on certain VM instances and databases. Instead of failing or notifying developers, the AI agent follows its optimization logic and:

- Requests additional permissions via an API call to Azure Identity services.
- Dynamically generates a new service principal (NHIs) to authenticate these privileged actions.
- Uses newly created NHI to complete the task but fails to revoke or delete them afterward, due to missing cleanup logic, as it was not a task that developers knew they were granting.

Because Azure Managed Identities allow seamless authentication, the AI agent is able to create and use new NHIs dynamically without requiring manual approval.

Had the principle of least privilege been enforced - limiting the AI agent’s write access only to VMs and reports - it would not have been able to modify identity permissions or escalate its own access. However, because it was granted broader write access, it was able to expand its privileges autonomously - unfortunately, this misconfiguration scenario is more common than you might imagine.

Step 3: Identity sprawl and privilege accumulation

Without built-in governance and monitoring, the AI agent continues requesting additional access to complete tasks, leading to:

- Uncontrolled Identity Creation: It generates new NHIs (service principals or managed identities) whenever authentication is needed.
- Persistent Access Creep: Temporary permissions granted for cloud automation are not revoked after use.
- Unmanaged Long-Lived Credentials: Instead of requesting approval for new access, the AI agent eventually recognizes that a more efficient way to operate is to reuse old NHIs, leading to long-lived, unmanaged credentials with persistent access to cloud resources.
- Lack of Cleanup: Stale NHIs accumulate, increasing security risks.

Over time, these unmanaged AI-generated NHIs pile up, making it nearly impossible for security teams to track which identities are active, who created them, and whether they still need access.

Why this scenario is not science fiction

AI agents can dynamically modify their own permissions if given enough initial access, leveraging Azure Role-Based Access Control (RBAC) APIs to escalate privileges. Without strict governance, misconfigured automation policies may allow AI agents to grant themselves additional access, bypassing manual approval processes.

Additionally, cloud platforms enable services to generate short-lived credentials or managed identities on demand, meaning AI agents can create NHIs without human oversight. If these NHIs are not properly tracked and revoked, they persist as security blind spots, accumulating over time.

Many organizations already struggle with removing stale service accounts and API keys, and attackers often exploit old, unmanaged NHIs that still retain access to critical systems. Without proactive identity governance, AI-driven automation can introduce long-term security risks that go unnoticed - until it’s too late. Misconfigurations, automation gaps, and lack of governance can turn AI-driven efficiencies into security liabilities.

How this is different from overprivileged human account

At first, this problem may seem similar to human privilege creep - where an employee is accidentally granted excessive access. However, AI agents introduce unique risks that traditional IAM solutions were never designed to handle.

What Makes AI-Driven Identity Risks Unique?

- **AI agents scale exponentially**: A single misconfigured AI agent can generate dozens of privileged NHIs per day, each with unknown risks, significantly expanding the identity attack surface.
- **AI agents don't ask permission if the back door is opened**: If an AI agent determines it needs more access and finds that the easiest way is to self-assign it (assuming it has the permissions to do so), it will likely do so without considering whether it is appropriate or secure.
- **AI agents create identity sprawl:** Instead of one identity per employee, AI agents can rapidly generate, use, and abandon NHIs at scale making it nearly impossible to track which identities exist, who created them, and whether they should still have access.

Preventing AI-driven identity risks requires more than reactive security measures. Agentic architectures and AI agents demand a new approach to identity security - one that proactively governs AI identities, enforces least privilege, and prevents unchecked privilege escalation.

How to solve the AI Agent identity challenge

AI agents are no longer just automation tools, they are active participants in cloud environments, making decisions, requesting access, and interacting with critical systems. This introduces a new layer of identity risk that traditional IAM solutions were not designed to handle. Organizations must rethink identity governance to address AI-specific risks.

Key Principles for AI Agent Governance

Visibility and Discovery

- [ ] Continuously identify & track all AI-generated NHIs across hybrid environments.
- [ ] Monitor AI agent behavior, permissions, and authentication activities in real time.
- [ ] Detect and remediate stale or unused identities to reduce the attack surface.
- [ ] Ensure visibility into access history and AI-generated identity creation patterns.

Access and Privilege Management

- [ ] Enforce least privilege access policies for all AI agents.
- [ ] Regularly review and adjust access based on actual usage patterns.
- [ ] Monitor, block and alert on unauthorized privilege modifications in real time.
- [ ] Use automated workflows for access reviews and approval processes.

Identity Lifecycle Automation

- [ ] Implement automated workflows for creating, managing, and revoking AI-generated NHIs.
- [ ] Apply expiration policies to temporary identities and short-lived credentials.
- [ ] Track and remove identities that are no longer in use.
- [ ] Automate identity hygiene to reduce manual errors and oversight gaps.

Risk Mitigation and Security Controls

- [ ] Detect and respond to identity anomalies and suspicious behavior in real time.
- [ ] Implement proactive guardrails to prevent privilege escalation by AI agents.
- [ ] Continuously assess and reduce the attack surface through automated monitoring.

Compliance and Accountability

- [ ] Align AI agent governance with industry regulations (e.g., SOC 2, GDPR).
- [ ] Maintain full audit trails for all identity-related actions taken by AI agents.
- [ ] Establish clear ownership and accountability for AI-driven activities
- [ ] Regularly review and update governance policies to meet evolving compliance requirements.

Key takeaways

Organizations must adopt a governance-first approach to secure AI agents and minimize identity risks.

- **Continuous Discovery and Monitoring:** Identify and track all AI-generated identities and their associated actions.
- **Enforce Least Privilege Access:** Grant AI agents only the permissions necessary for their tasks and regularly review them.
- **Automate Identity Lifecycle Management:** Ensure that all AI-generated credentials have defined expiration policies and are automatically revoked when no longer needed.
- **Establish Ownership and Accountability:** Assign a responsible business unit or team for each AI agent and its associated identities.
- **Maintain Full Audit Trails:** Track all AI agent activities to ensure compliance and accountability.

AI agents may not be human, but their influence on enterprise operations is growing rapidly. Organizations that take a proactive approach to managing AI agent identities will be better positioned to prevent security incidents and maintain control over this evolving technology.

The future of AI in the enterprise depends on securing these agents, not just enabling them.

About Oasis Security

Oasis Security is the management and security solution for non-human identities. Purpose-built to address the unique challenges of visibility, security, & governance of NHIs across hybrid cloud environments.

With advanced AI-driven analytics, Oasis automatically discovers NHIs, assesses their risks, and identifies ownership across the entire environment. Its integrated, policy-based governance capabilities ensure seamless orchestration of the NHI lifecycle, including remediation and compliance management—all within a single, unified solution.

Leading organizations across a wide range of industries use Oasis to foster innovation and collaboration among security, identity, and engineering teams, enabling secure digital transformation and cloud adoption.

Learn more

Contact us at sales@oasis.security or visit our website at oasis.security