E-Book

The AI Agent Era: Rethinking Identity, Trust, and Control

Contents

The new identity frontier

Traditional IAM is built for human speed.
AI moves at machine speed.

As organizations transition from static chatbots to autonomous agents, the fundamental nature of identity is changing. These agents don't just provide information; they execute multi-step tasks across complex environments, creating the potential for massive visibility gaps where unauthorized actions can occur before a security team can even detect them.

This eBook serves as a compilation of key highlights from our comprehensive 7-part thought leadership series on securing the AI-powered organization. Beyond summarizing those core discussions, this guide provides deeper strategic insights into the architectural shifts required for autonomous agency.

91%

of organizations are already using AI agents in production.

10%

of organizations have a well-developed strategy for managing non-human identities.

+50%

cite AI agent governance and compliance as a top concern.

source

IAM at agent velocity

As AI agents move at machine speed, executing up to 6,000 operations per minute, traditional human-centric authorization models collapse. Security must shift from manual, consent-based approvals to automated, runtime enforcement to prevent "panicked" agents from causing catastrophic data loss in seconds.

Recommendations

- **Policy-driven rules:** Implement automated rules that scale to agent velocity, so that security keeps pace with machine-speed execution.
- **Ephemeral credentials:** Utilize credentials that expire in minutes instead of persisting indefinitely, drastically shrinking the window of opportunity for attackers.
- **Relationship-based access:** Enable millisecond authorization checks using fine-grained, relationship-based access control.
- **Continuous evaluation:** Move away from "grant and forget" permissions toward reassessing every single operation an agent performs.

Can your current security stack intercept a rogue agent executing 100 commands per second before it deletes your database? If you are still relying on human sign-offs for agent actions, you aren't just slowing down innovation, you’re ignoring a machine-speed breach.

Read: AI Security: IAM at Agent Velocity

When authorization outlives intent

"Authorization drift" occurs when digital keys issued for a specific task remain active for months after the job is done. In the world of AI agents, these dormant tokens are ticking time bombs, allowing attackers to hijack legitimate SaaS-to-SaaS connections without ever needing to crack a password.

Recommendations

- **Durable delegated identity:** Every AI agent must have its own identity, separate from users, that is governed, auditable, and traceable.
- **Continuously renewable authorization:** Adjust access automatically as the task, user, or environment changes to maintain a zero standing privilege posture.
- **Instant cross-system deprovisioning:** Enforce shared-signal revocation so that a credential revoked in one application is instantly invalidated everywhere.
- **Real-time intent validation:** Re-check every action against current policies at the moment they happen, not just when credentials were originally issued.

Does your organization have a formal, automated process for revoking agent credentials the moment a task is completed? If your tokens outlive their intent, you are leaving your back door wide open for "silent" breaches.

Read: AI Agent Security: When Authorization Outlives Intent

Your AI security and compliance action plan

Ready to take control of your AI agent ecosystem? Access the complete AI Identity Security Compliance Checklist to guide your next steps.

Get the checklist to learn more

Agent cross-system trust

When AI agents cross organizational boundaries to access independent systems, they often lose their "memory" of constraints. Because most identity providers validate tokens in isolation, a compromise in one trust domain can ripple unchecked across hundreds of others.

Recommendations

- **Verifiable delegation:** Implement cryptographic proof of delegation that explicitly differentiates between human and agent identities as they move between systems.
- **Portable constraints:** Maintain that security constraints (like "read-only") travel with the token across trust domains so they aren't stripped away during a handoff.
- **Coordinated revocation:** Adopt federated risk signaling (like IPSIE) so that real-time security alerts are shared between different service providers.
- **Retrieval-time controls:** Use fine-grained authorization to validate access at the exact moment an agent calls an API, regardless of which domain the request originated from.

When an agent crosses into a partner's system, who is vouching for its current safety? If your trust model is static and decentralized, you have no shared defense against a cross-domain token hijack.

Read: AI Security: Agent Cross-System Trust

Securing AI agent delegation

Recursive delegation—where agents spawn sub-agents to handle specialized tasks—creates a "Russian nesting doll" of security risks. Without strict lineage and scope attenuation, a single malicious prompt can trigger "Agent Session Smuggling," allowing a sub-agent to execute unauthorized actions invisibly.

Recommendations

- **Out-of-band verification:** Use push notifications or separate UIs for high-consequence operations to bypass the agent's primary chat channel.
- **Context grounding:** Anchor every agent session to its original task and continuously flag "semantic drift" if the agent's behavior begins to wander from the intended goal.
- **Verified identity and capabilities:** Enforce a zero-trust architecture where agents must present cryptographic credentials to verify their identity and specific permissions before interacting with any system resource.
- **User visibility:** Mitigate the risk of smuggled instructions by providing radical transparency, surfacing all tool calls, background reasoning, and execution logs to the user in real-time.

Can you cryptographically prove the lineage of every action taken by a sub-agent three hops deep into a workflow? Without verifiable delegation chains, your multi-agent ecosystem is a high-leverage target for lateral movement.

Read: Agent Security: Delegation Chain

Closing the authorization gap

AI agents often retrieve data using an executive's high-level permissions but output that information into shared workspaces like Slack or Teams. This "authorization gap" allows sensitive data, such as executive compensation or board materials, to be inadvertently broadcast to unauthorized recipients.

Recommendations

- **Audience-aware authorization:** Compute the "permission intersection" in real time so that an agent only retrieves data that everyone in the current workspace is authorized to see.
- **Scoped retrieval:** Shift from filtering data after it is fetched to scoping the agent's credentials before retrieval, so that sensitive files are never even accessed.
- **Relationship-based access:** Move beyond static roles to a relationship-based model that understands who is in which channel and what their current permissions are.
- **Continuous policy sync:** Integrate identity governance with the authorization engine so that the "permission graph" remains accurate as users join or leave shared workspaces.

For every AI agent deployed in a shared workspace, can you demonstrate that its output is restricted to the "lowest common denominator" of permissions in the room? If your agents ignore the audience, they are your biggest internal data leak risk.

Read: AI Agent Authorization Gap

Cyber-physical safety

As AI agents move into physical industries, like healthcare and manufacturing, authorization errors shift from data leaks to safety hazards. Security must now prevent digital agents from causing physical damage.

Recommendations

- **End-to-end traceability:** Implement Cross-App Access (XAA) with delegation tokens to help attribute every automated action to both the specific agent and the originating user.
- **On-demand credentialing:** Utilize a Token Vault to replace static, long-lived credentials with scoped, short-lived tokens that are retrieved only at the moment of execution.
- **Human-in-the-loop (HITL) verification:** Use Client-Initiated Backchannel Authentication (CIBA) to require explicit human approval whenever an agent attempts to perform high-consequence or "out-of-envelope" operations.
- **Real-time policy enforcement:** Shift to fine-grained authorization that evaluates safety constraints and operational limits at the exact moment of the decision, rather than relying on static roles.

Can your team clearly define the authorization envelope and active credentials for every agent accessing critical systems? If you cannot articulate exactly what an agent is permitted to do and why, that lack of visibility is your primary attack surface.

Read: AI Agents: Cyber-Physical IAM Safety

Unified identity and authorization for autonomous trust

As AI agents operate autonomously, traditional identity and authorization models break down. Systems designed to grant access to human users cannot reliably govern delegation, indirect actions, or machine-driven decision-making. Establishing autonomous trust requires treating identity and authorization as a unified control layer that defines what agents are allowed to do—and where those boundaries end.

Recommendations

- **Represent agents as distinct identities:** Treat AI agents as their own identities rather than as extensions of users or applications.
- **Unify identity and authorization decisions:** Unify identity and authorization decisions: Avoid fragmented controls by enforcing consistent access decisions across agent actions.
- **Account for delegation explicitly:** Design authorization models that recognize when agents act on behalf of users, systems, or other agents.
- **Reduce over-broad access:** Limit standing permissions that allow agents to operate beyond their intended scope.
- **Use identity to define boundaries:** Apply identity and authorization as mechanisms for constraining agent behavior, not just observing it.

When identity and authorization operate as a unified system, autonomous trust becomes enforceable as agents act and delegate.

Read: Identity and Authorization as the Operating System for Autonomous Trust

How Okta can help

Okta helps bring the identity security fabric to life as the unified control plane by bridging the gap between human intent and machine-speed execution. It provides a centralized layer for real-time, policy-driven enforcement, so that every agent action is governed, traceable, and secure across any trust domain. By making identity the core of the infrastructure, the fabric allows organizations to scale AI agents with the confidence that safety and authorization are hardwired into every automated workflow.