ConductorOne

A CISO's Guide to Agentic AI

Why agentic AI matters now

The rise of agentic AI represents a seismic shift in the security landscape. Unlike traditional AI models that analyze data or generate content, agentic AI systems are capable of taking autonomous action across systems, often without direct human oversight. These agents can trigger workflows, access applications, interact with other agents, and make real-time decisions—transforming how organizations operate.

For CISOs, the question is no longer if agentic AI will impact security. It's how quickly they can adapt. Traditional identity and access management (IAM) models simply don't scale to meet the demands of thousands of short-lived, task-specific, autonomous AI identities. This guide is designed to help security leaders define, evaluate, and act on this rapidly evolving threat and opportunity.

Understanding agentic AI

Agentic AI refers to autonomous systems that act on behalf of users or organizations to complete tasks. These agents do more than analyze or recommend; they make their own decisions and execute without human direction. This represents a fundamental shift in how automation is deployed within the enterprise.

These AI agents can be:

- → **Ephemeral:** They may exist for seconds or minutes, completing a specific task before disappearing.
- → **Autonomous:** They operate without constant human intervention, evaluating context and taking action dynamically.
- → **Task-based:** They require granular, scoped permissions tied to specific functions and should never be broadly privileged.

Understanding the architecture and behavior of these agents is foundational for security leaders seeking to govern them effectively. Without visibility into what these agents are doing and what they have access to, organizations face serious risk.

Types of agentic AI

To govern agentic AI effectively, CISOs need to understand the landscape. AI agents generally fall into three categories:

1. Company AI agents

These are system-level agents embedded in core business platforms (for example, within Salesforce, GitHub, or ServiceNow). They execute workflows on behalf of the business and are typically deployed organization-wide.

- → Operate like highly intelligent service accounts
- → Execute predefined workflows (e.g., lead scoring, code reviews)
- → Introduce risk if over-permissioned or if workflows go unmonitored

2. Employee AI agents

These agents act as digital assistants for individual users. They span tools, automate repetitive tasks, summarize information, and enhance productivity.

- → Inherit permissions from the employee
- → Increase productivity, but risk acting beyond the user's intention
- → Require a new paradigm for delegated authority and visibility

3. Agent-to-agent interactions

The most complex and least understood category, these agents coordinate and communicate autonomously to complete multi-step tasks across systems.

- → Represent a new frontier in automation
- → Raise serious questions around auditability and authorization
- → May chain together actions that exceed the original intent

Understanding these types is critical, as each carries different risk profiles and governance needs.

The scale problem: why identity is breaking

Traditional IAM systems were built for a human workforce. In that model, an employee is onboarded, granted access over time, and periodically reviewed. AI agents break that mold entirely:

→ **Lifespans are short**: Agents may exist for mere seconds

→ **Access needs are highly dynamic**: Permissions must shift in real-time

→ **Manual approvals are infeasible**: Human-in-the-loop slows down machine-native processes

As a result, identity systems are being stretched beyond their limits. Manual access reviews and traditional role-based access control (RBAC) can't keep pace. What began as a manageable human-centric problem has become an overwhelming challenge, especially when AI can create and interact with other AI.

A single AI task could generate multiple identities across systems. Without proper scoping and oversight, this sprawl creates massive attack surfaces, shadow access, and critical blind spots.

The trust problem: rethinking assurance

Trust in an AI-native world is more than just verifying credentials. CISOs must develop a multi-dimensional model of trust that combines identity with:

→ **Behavioral analytics**: What is the agent doing, and is that behavior expected?

→ **Data provenance**: Are inputs and decisions based on trustworthy data sources?

→ **Access scoping**: Does the agent have permission to do this specific action?

→ **Real-time observability**: Can the agent's actions be monitored, traced, and revoked?

→ **Governance alignment**: Are actions and access decisions in line with organizational policy?

→ **Auditability**: How can we trust but verify?

In addition to technical trust, agentic AI raises ethical and compliance concerns. When agents operate autonomously, they may access or share data in unintended ways. The principle of least privilege becomes harder to enforce, and the consequences of misbehavior multiply rapidly.

What are the biggest risks CISOs should think about?

1. **Expanded Attack Surface**
   Agentic AI systems may have access to multiple systems and sensitive data. If compromised, they can act independently, amplifying the potential impact of a breach.
2. **Autonomy Equals Risk**
   Without proper guardrails, agentic AI might take unintended or unsafe actions (e.g., overly broad access approvals or incorrect threat remediations).
3. **Governance and Compliance Challenges**
   Ensuring that autonomous actions meet regulatory, audit, and ethical standards becomes more complex. CISOs need transparency and traceability in AI decisions.

4. **Insider Threat Amplification**
   A malicious actor could manipulate an AI agent (e.g., prompt injection or logic corruption), effectively turning it into a high-speed, persistent insider threat.
5. **AI Supply Chain Risks**
   Agentic capabilities may come from third-party tools or libraries. Ensuring their integrity, update cadence, and security posture is critical.

How to govern agentic AI today

Despite the complexity, CISOs can take decisive action now. Leading organizations are embracing a multi-pronged approach that combines new architectures, updated processes, and strategic oversight.

1. Develop a clear strategy

The first step in adopting agentic AI is to develop a clear, evolving strategy. CISOs should work closely with the business to identify and prioritize beneficial use cases while restricting high-risk or misaligned ones. When implemented thoughtfully, agents can accelerate business operations by enhancing efficiency, delivering faster and more detailed insights, and scaling human capabilities.

2. Discover and classify agents

Discover the agents already in use in your organization. Map all AI agents operating in your environment:

→ Where are agents deployed?

→ What systems do they touch?

→ What decisions are they making?

This inventory is foundational to building governance controls. After you discover and inventory the agents in use, create an easy to use process for your organization to bring new agents in. This allows you to be proactive, rather than just reactive.

3. Ensure agents have identities and support ephemeral credentialing

Agents must have identities to enable secure authentication, policy enforcement, and activity tracking. With identities in place, agent access should rely on ephemeral credentialing rather than static credentials, which pose significant risk. Implement:

→ One-time, task-scoped credentials

→ Dynamic validation tied to context (e.g., task type, source system)

This approach minimizes standing access and ensures agent actions are constrained, intentional, and auditable.

3. Ensure you have the ability to support task-based authorization

RBAC falls short in agentic environments. Use:

→ Context-aware policies that adapt dynamically

→ Access scopes tied to workflows, not roles

→ Continuous evaluation of permissions

4. Implement AI-native identity governance

Look for or build systems with:

→ Support for ephemeral and high-volume identity creation

→ Automated deprovisioning workflows

→ Integration with model context protocols (MCPs)

→ Purpose-built audit and monitoring pipelines

→ Support agent-to-agent protocols

5. Build auditability, observability, and telemetry

Without monitoring, governance is toothless:

→ Monitor for drift and misuse

→ Log agent decisions, success/failure rates, and permission usage

→ Enable root cause analysis when things go wrong

6. Keep a human in the loop

Even with autonomous agents, human oversight and layered controls are essential for security. It's not just about approvals, it's about validating that intended actions actually occurred. To build depth and resilience:

→ Require approvals for access to sensitive data or high-impact systems

→ Use layered controls like dual-agent validation or approval queues

→ Implement kill switches, rollback options, and post-action verification mechanisms

Security depends on depth and validation layers: ensuring that what was supposed to happen did happen.

Executive alignment and cross-functional collaboration

CISOs are not in this alone. As executive buy-in accelerates AI adoption, security teams must collaborate across engineering, data, legal, and compliance to:

→ Establish AI adoption frameworks

→ Define risk thresholds and control expectations

→ Balance speed with responsibility

To formalize this effort, consider creating an AI oversight committee. This cross-functional group can lead discovery, define the golden path for agent adoption, and enforce governance over time.

It's also critical to align with procurement and vendor risk teams, as third-party tools increasingly ship with embedded AI agents. Agent-native products can introduce opaque access models and expand supply chain risk.

The path forward

Agentic AI is here. It is redefining identity, reshaping access governance, and challenging the foundations of enterprise security. Organizations that wait will be overwhelmed. Those who act now have a chance to set the standard.

Agentic AI is not just another automation layer. It's a new class of actor in your environment. You must treat it with the same scrutiny you apply to human users, privileged accounts, and critical systems. That means CISOs must:

→ Replace human-centric identity models with agent-native frameworks

→ Automate governance at the speed of machine decision-making

→ Continuously adapt policies to reflect an evolving ecosystem

→ Treat every AI agent as a first-class identity with explicit controls

There is no "done" in this world. Governance must evolve as fast as the threats.

Summary: CISO checklist for agentic AI readiness

- **Develop a strategy:** Identify and prioritize beneficial use cases while restricting high-risk or misaligned ones
- **Inventory all AI agents:** Document agents and their access scopes
- **Implement ephemeral, dynamic credentials:** Replace static credentials with short-lived, context-aware access
- **Replace RBAC with task-based authorization:** Use policies tied to workflows and real-time context
- **Deploy AI-native identity governance:** Support high-volume identity creation, automation, and model context protocols
- **Monitor agent activity:** Enable real-time observability, logging, and anomaly detection
- **Require human oversight:** Add approvals, validation, and fail-safes for high-impact workflows
- **Continually evaluate and adapt:** Refine access models to scale with evolving environments