

Securing Agentic AI: Identity as the Emerging Foundation for Defense

This report draws on multiple sources including research and analysis from McKinsey & Company. Palo Alto Networks is responsible for the conclusions and recommendations reflected in this paper.

Table of Contents

- How Agentic AI Differs from Generative AI3
- Agentic AI Adoption Is Accelerating3
- Six Emerging Types of AI Agents4
- Amplification and Mitigation of Agentic AI Risks5
- Shifts in the Security Stack and the Increasingly Important Role of Identity7
- Governance Structures Evolve to Address Agentic AI Risk8
- Emerging Market Dynamics, Needs, and Areas of Opportunity8
- Conclusion9
- About Palo Alto Networks9

As enterprises continue deploying autonomous AI agents at scale, cybersecurity leaders are confronting new security and governance challenges. Traditional security controls for managing human and machine identities need to be enhanced to govern autonomous, decision-making systems that operate with elevated permissions and access. Identity is increasingly emerging as the control plane for securing these new digital workers.

This paper draws on multiple sources including a survey of over 100 CISOs across enterprise financial services and software companies and insights from cybersecurity, technology, and AI leaders. It defines agentic AI, examines how adoption is unfolding, outlines the emerging types of agents, the control frameworks needed to manage them safely, and the evolving role of identity security for agentic AI.

Identity management is being defined as the primary control plane through which agents are governed, credentialed, and enabled.

How Agentic AI Differs from Generative AI

An AI agent is defined as an autonomous software program designed to perceive its environment, make decisions, access resources, and invoke tools and APIs to achieve predetermined goals, often using large language models (LLMs) and machine learning.¹

Unlike generative AI, which produces output when prompted, agentic AI executes tasks independently, linking models, data, and systems to achieve defined goals. This shift from content generation to autonomous execution amplifies existing risks and introduces new risks that traditional security frameworks are not sufficiently designed to manage.

Agentic AI Adoption Is Accelerating

Organizations are embedding autonomous AI agents to improve business processes and deliver measurable impact. Agentic AI is unlocking new value across functions, ranging from front-office activities such as marketing and customer service to supporting security operations.² Building on the economic potential of generative AI, which McKinsey estimates could add \$2.6 trillion to \$4.4 trillion annually to the world economy,³ agentic AI represents the next stage of adoption as organizations move from content generation to autonomous execution.

Four in 10 enterprise financial institutions and software companies surveyed in North America and Europe already report having agentic AI in production, and three-quarters expect to do so within the next three years, meaning agents will no longer be confined to pilot programs but embedded in core business systems and everyday operations.⁴

1. "What Is Agentic AI vs AI Agents," CyberArk, accessed April 12, 2026.

2. "Digital Workers Have Arrived in Banking," The Wall Street Journal, June 30, 2025. The article describes BNY Mellon's deployment of its first "agentic employee," an application security tester designed to automate vulnerability assessment and compliance validation.

3. *The economic potential of generative AI: The next productivity frontier*, McKinsey & Company, June 14, 2023.

4. Findings are based on a September 2025 survey of 104 CISOs and equivalent security leaders across North America (56%) and Europe (44%), covering large enterprise financial services (54%) and software/SaaS companies (46%). The survey was supplemented by interviews with cybersecurity and AI practitioners.

Six Emerging Types of AI Agents

AI agents can be categorized into one of six archetypes with differing levels of autonomy (see Exhibit 1).

Agent Type	Description
Knowledge/research	Provides information and analysis across domains (e.g., summarize market research or regulatory updates)
Task	Performs simple, discrete actions triggered by workflows (e.g., schedule meetings or send compliance reminders)
Workflow	Executes multi-step operational processes across systems (e.g., orchestrate client onboarding or prepare standard reports)
Goal/planning	Builds plans and determines execution paths to meet defined objectives (e.g., create staffing schedules or investment strategies)
Orchestrator	Coordinates multiple agents or systems to optimize execution (e.g., oversee product release cycles or manage system migrations)
Personable agents	Acts as virtual teammates with human-like interaction and decision-making (e.g., support customer service or deliver internal training)

Exhibit 1. Taxonomy of agent types

Lower autonomy agents such as task and workflow agents are scaling the most, while higher autonomy agents such as goal planning agents are cautiously following. This adoption curve signals a transition from experimentation to enterprise-wide reliance on AI agents over the next three years (Exhibit 2).⁵

Adoption Levels of Agent Types Over Time

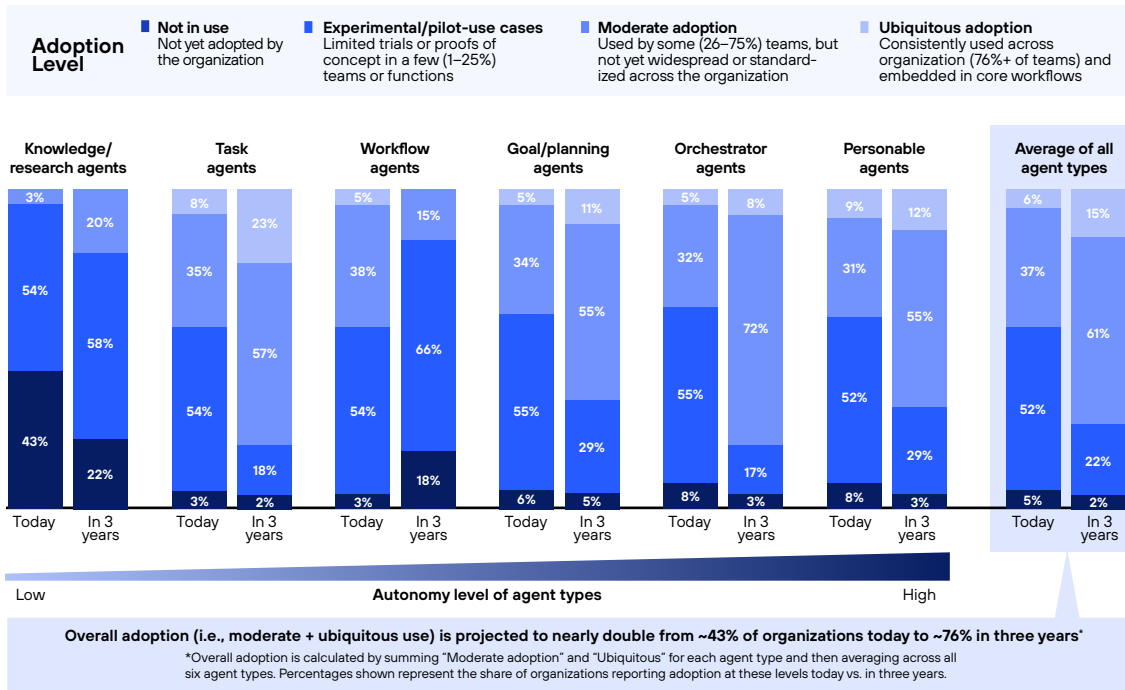


Exhibit 2. Agent AI adoption

5. Findings are based on a September 2025 survey of 104 CISOs and equivalent security leaders across North America (56%) and Europe (44%), covering large enterprise financial services (54%) and software/SaaS companies (46%). The survey was supplemented by interviews with cybersecurity and AI practitioners.

Not all agents pose the same level of risk. While risk can vary by implementation, research agents that retrieve information typically present lower exposure than orchestrator agents, which can initiate or coordinate actions across systems.

To apply the right level of security and oversight, organizations should establish a taxonomy of agent types and their corresponding risk levels, as shown in Exhibit 3.

		Most frequently prioritized risks (per agent type) (% of organizations indicating it as a top risk, N=104)		
Autonomy	Agent Type			
High 	Personable agents Act as virtual teammates with human-like qualities	Synthetic identity risk (33%)	Cross-agent task escalation (26%)	Untraceable data leakage (22%)
	Orchestrator Coordinate multiple agents or systems	Cross-agent task escalation (43%)	Untraceable data leakage (15%)	Data corruption propagation (15%)
	Goal/planning Build plans and determine execution paths	Cross-agent task escalation (28%)	Data corruption propagation (25%)	Untraceable data leakage (19%)
	Workflow Execute multi-step processes across systems	Cross-agent task escalation (38%)	Untraceable data leakage (20%)	Chained vulnerabilities (16%)
	Task Perform simple, discrete tasks triggered by workflows or prompts	Cross-agent task escalation (33%)	Untraceable data leakage (23%)	Synthetic identity risk (20%)
	Knowledge/research Provide information and analysis	Untraceable data leakage (37%)	Cross-agent task escalation (25%)	Data corruption propagation (24%)
Low				

Source: Industry Survey on agentic AI (n=104), Sept. 2025. Q1.12 asked for each agent type which risks are most important to manage (ranked by respondents).

Exhibit 3. Agentic AI risks

Amplification and Mitigation of Agentic AI Risks

Agentic AI both amplifies existing cybersecurity risks and introduces new ones. Familiar risks such as data leakage or access misuse become harder to monitor when agents exchange information across systems without sufficient oversight. New risks arise when agents impersonate one another or exploit trust mechanisms to trigger unauthorized actions. The autonomous nature of these systems (especially at scale) creates failure modes that are difficult to detect and contain, such as invisible inter-agent data flows (see Exhibit 4).

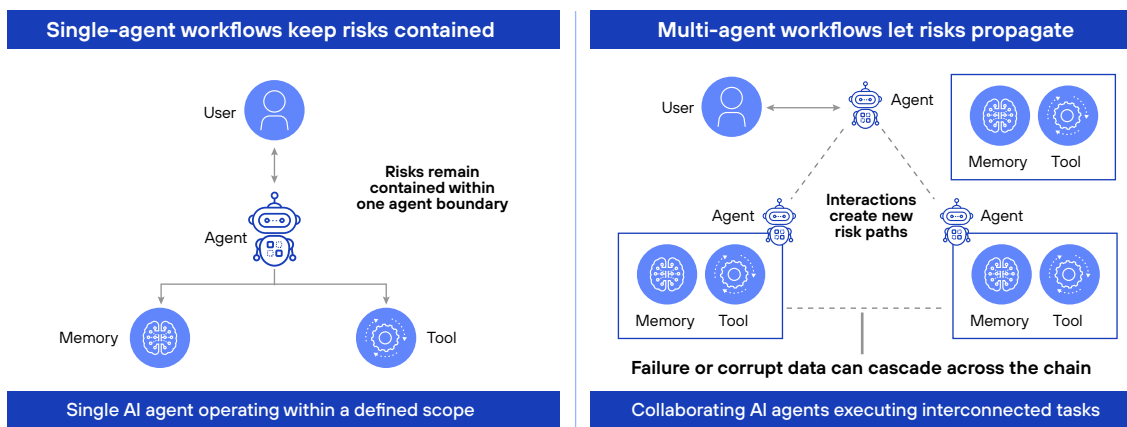


Exhibit 4. Single vs. multi-agent risks

Two-thirds of CISOs surveyed in financial services and software rank agentic AI among their top three cybersecurity risks—and more than one-third name it as their top concern, ahead of ransomware and supply-chain threats.⁶

Tracing the full agentic AI lifecycle, from how agents are built and deployed to how they act and interact, shows how autonomy reshapes risk. Once agents can make decisions, delegate tasks, and share information, familiar issues evolve into systemic exposures that propagate across workflows. The resulting failure modes can be grouped into distinct categories of agentic AI risk (Exhibit 5).






Novel Agentic AI Risk (traditional anchor)	Description	Example
 Cross-agent task escalation (privilege escalation)	Agents hand off tasks or permissions in ways that escalate beyond intended scope	A rogue scheduling agent impersonates a physician and convinces a clinical agent to release patient records
 Untraceable data leakage (data exfiltration)	Sensitive data flows between agents/tools without logs or auditability	A customer support agent quietly shares account details with an external "fraud detection" agent without logging the transfer
 Synthetic identity risk (identity spoofing)	Forged or cloned agent IDs used to bypass trust or accountability	An attacker forges a claims-processing agent ID to file fake insurance claims that the system accepts as legitimate
 Chained vulnerabilities (supply chain flaws)	Small weaknesses across multiple agents/tools combine into systemic exploits	DevOps orchestrator agent misinterprets a deployment policy and spins up a container with admin/root privileges
 Propagation of corrupted data (data corruption)	Bad or poisoned data spreads automatically through agent workflows	A mislabeled clinical trial dataset poisons a labeling agent, whose output corrupts downstream efficacy and regulatory reporting agents

Exhibit 5. Agentic AI risk taxonomy

To address the amplified and new risks introduced by agentic AI, traditional security controls provide only partial protection. Identity and access management, role-based access controls, and data loss prevention remain foundational, but they alone cannot fully mitigate the risks of autonomous systems. These controls were designed for static users and predefined access patterns, not for systems that can act, decide, and interact independently.

Organizations will need to extend their control frameworks with new capabilities purpose-built for autonomy across four priority areas:

- **Identity and access controls** to manage how agents are created, credentialed, and retired (e.g., agent registries, dynamic just-in-time access permissions)
- **Visibility and monitoring controls** to track and log agent activity so behaviors remain traceable and auditable (e.g., continuous activity logs, agent anomaly dashboards)
- **Containment and recovery controls** to isolate or shut down compromised agents (e.g., built-in kill switches, automatic rollback to safe system states)
- **Data assurance controls** to validate where data comes from, who or what can access it, and how it is used to prevent errors or manipulation as agents learn and act (e.g., source validation tools, access approvals for sensitive information)

6. Findings are based on a September 2025 survey of 104 CISOs and equivalent security leaders across North America (56%) and Europe (44%), covering large enterprise financial services (54%) and software/SaaS companies (46%). The survey was supplemented by interviews with cybersecurity and AI practitioners.

While overall adoption of AI agents is expected to reach 76% in the next three years, fewer than 1 in 10 of organizations surveyed have deployed risk registries and dynamic authorization agent security controls at scale.⁷

Yet, more than 60% of CISOs rank registries and dynamic authorization as critical for securing agentic AI.⁸

Without faster adoption, agentic autonomy could outpace containment measures within the same time frame.

Shifts in the Security Stack and the Increasingly Important Role of Identity

Agentic AI is reshaping how cybersecurity resources are allocated. While governance and data security are gaining importance, driven by regulation and the need for auditability, identity is emerging as the anchor control plane for managing autonomous agents.

Identity management is both absorbing a larger share of cybersecurity budgets and being redefined as the primary layer through which agents are governed, credentialed, and contained. As agentic AI proliferates, identity's role expands from managing human access to orchestrating the full lifecycle of agent identities.

Agentic identities combine traits of human and machine identities but add new dimensions such as contextual autonomy, delegated authority, and non-deterministic behavior. These require new governance mechanisms and control extensions to maintain accountability and containment.

Within three years, identity is expected to represent nearly a quarter of total cybersecurity spend, which is a greater share than network and application security combined.⁹

Identity is evolving from authentication to orchestration. New capabilities such as kill switches, registries, and context-aware provisioning enable continuous oversight of agent behavior, while existing identity security technologies must scale to manage a surge of AI agents, which can have privileged machine identities.

Governance and data are becoming co-control planes. Driven by regulations such as the EU AI Act and DORA, they are absorbing close to one-third of budgets as organizations strengthen lineage tracking, encryption, and auditability to demonstrate the trustworthiness of AI-enabled decisions.¹⁰

Perimeter-based controls are receding in importance. As infrastructure becomes software-defined and identity-anchored, data-driven architecture becomes the default defense model.

7. Agent registries maintain an authoritative record of deployed AI agents, including ownership, permissions, and lifecycle status. Dynamic authorization (referred to as fine-grained or policy-based access control) issues just-in-time, context-aware credentials that adjust automatically based on task, privilege, or risk level.

8. Findings are based on an internal September 2025 survey of 104 CISOs and equivalent security leaders across North America (56%) and Europe (44%), covering large enterprise financial services (54%) and software/SaaS companies (46%). The survey was supplemented by interviews with cybersecurity and AI practitioners.

9. Ibid.

10. Ibid.

Governance Structures Evolve to Address Agentic AI Risk

Beyond the technologies and controls required to mitigate the risk of agentic AI, effective governance is essential. Today, one-third of organizations assign responsibility to a Chief AI Officer or Chief Data Officer, another third to CISOs, and the rest across CIOs, CTOs, or joint committees.¹¹

Organizations are adopting one of three governance archetypes to manage agentic AI risk, with the right approach depending on an organization's size, regulatory environment, and role of AI in its business model.

- **Centralized governance**, such as under a Chief AI Officer or an AI Risk Committee, concentrates accountability under a single leader or committee, ensuring consistent standards and regulatory alignment. This model is most effective when AI is core to strategy or revenue, or when board-level visibility is required, but it can slow execution if additional review layers are added.
- **Functional governance**, led by roles such as a Chief Risk Officer, Chief Legal Officer, or Chief Information Security Officer, embeds responsibility within existing functions, enabling faster adoption and alignment with established processes. This approach fits organizations with mature risk-management frameworks but can still leave cross-functional gaps, since emerging AI risks often cut across data, technology, and operational boundaries.
- **Federated governance**, typically business unit- or regional-entity-led, allows flexibility and speed where AI risks vary across the organization. This model suits diversified or digital-first organizations but can result in inconsistent standards and limited enterprise-wide visibility if not coordinated centrally.

Emerging Market Dynamics, Needs, and Areas of Opportunity

To securely adopt agentic AI agents at scale and obtain the corresponding benefits, organizations should consider five actions related to risks, adoption, governance, accountability, and entry points.

- **A clear taxonomy for agentic AI and its risks is needed.** Early work is underway, including efforts from OWASP (Top 10 for LLMs) and NIST (AI control overlays for SP 800-53),¹² but industry alignment on an agentic AI risk taxonomy remains in progress.
- **Adoption curve expectations need to be managed.** To realize long-term savings and efficiency gains, organizations must first expect upfront investment in control design, new governance processes, and training.
- **Identity and session governance must extend to autonomous AI systems.** As AI agents proliferate, identity is becoming the new anchor of trust, and organizations require scalable ways to manage access and contain agentic threats.
- **Accountability for agentic AI risk needs careful coordination.** Ownership of agentic AI security varies, creating uncertainty and inconsistencies in who leads investment and oversight; models and governance structures must align risk, technology, and compliance across security, AI, and business initiatives.
- **Buyers need clear, structured entry points.** Executives see potential in agentic AI and security leaders seek practical first steps to manage risk; teams need pragmatic guidance, including an agentic AI taxonomy that maps to key risks and control families.

11. Findings are based on an internal September 2025 survey of 104 CISOs and equivalent security leaders across North America (56%) and Europe (44%), covering large enterprise financial services (54%) and software/SaaS companies (46%). The survey was supplemented by interviews with cybersecurity and AI practitioners.

12. "Top 10 for Large Language Model Applications," OWASP, November 18, 2024; *SP 800-53 Control Overlays for Securing AI Systems Concept Paper*, National Institute of Standards and Technology, August 2025.

Conclusion

Agentic AI represents a transformative opportunity for business impact, but these opportunities will not be realized overnight—or without careful balance between automation and control. The priorities are the speed of adoption and the thoughtful design and deployment of guardrails that allow autonomy to operate within accountable boundaries.

Before agents can operate safely at scale, enterprises must first build the right foundations for trust (e.g., a clear risk taxonomy, resilient infrastructure, appropriate controls, and mechanisms to train and validate agent behavior). As these systems advance, and mimic more human behaviors at machine speed, agents will often require elevated permissions and access to perform tasks and meet goals.

However, all that capability equals risk. To safeguard AI agents, teams will require a combination of traditional and modern security controls, including a comprehensive identity security strategy for their AI agents. This, in turn, can help accelerate the deployment of AI agent registries, containment policies, and cross-functional governance. The result will better foster trust and resilience. Those that delay, however, will see autonomy outpacing oversight, a strategic vulnerability in an era likely to be increasingly defined by intelligent, non-deterministic agents.

Building the right foundations now, through taxonomy, governance, and controls, will determine which organizations can scale agentic AI without materially increasing systemic risk.

To explore all the ways Palo Alto Networks Idira™ can secure the identities across your organization, visit www.paloaltonetworks.com/idira.

About Palo Alto Networks

Palo Alto Networks (NASDAQ: PANW), the global AI cybersecurity leader, protects our digital way of life with a comprehensive portfolio of cybersecurity solutions and platforms across Network, Cloud, Security Operations, AI, and Identity. Trusted by more than 70,000 customers and powered by Unit 42® threat intelligence, our AI-driven platforms eliminate complexity, empowering enterprises to modernize with confidence and securing the speed of innovation. Explore the future of security at www.paloaltonetworks.com.



3000 Tannery Way
Santa Clara, CA 95054
Main: +1.408.753.4000
Sales: +1.866.320.4788
Support: +1.866.898.9087
www.paloaltonetworks.com

© 2026 Palo Alto Networks, Inc. A list of our trademarks in the United States and other jurisdictions can be found at <https://www.paloaltonetworks.com/company/trademarks.html>. All other marks mentioned herein may be trademarks of their respective companies.
idira_wp_idira_wp_securing-agentic-ai_042126