ARKOSE LABS

The 2026 Agentic AI Security Report

Enterprises Are Not Ready for the Next Wave of Agentic AI Attacks

Contents

Executive Summary

Agentic AI is now operating inside enterprise systems, while governance, visibility, and accountability are still catching up. This report is based on a global survey of 300 enterprise leaders across security, fraud, identity, and AI functions, including organizations with more than $1B in annual revenue. Respondents represent major technology platforms, global financial institutions, and large-scale enterprises, including many operating at Fortune 100 scale.

The findings reflect a shift in how enterprise risk is taking shape. AI agents are becoming active participants in operational workflows, using legitimate credentials and interacting across systems in ways that closely resemble trusted activity.

This introduces new challenges for security and fraud teams. Many existing models were designed around human behavior and external threats. Autonomous systems operate continuously across services and generate activity that is harder to isolate and interpret. As a result, understanding how actions occur and connect across systems is becoming central to investigation and response.

Enterprises are moving into a compressed window where deployment is accelerating faster than the controls required to manage it. Organizations that respond effectively will strengthen identity governance, improve attribution capabilities, and establish clear ownership and enforcement across teams.

Key Insights

Enterprise leaders expect near-term impact.

Nearly all **(97%)** of enterprise leaders expect a material AI-agent–driven security or fraud incident within 12 months, with almost half **(49%)** anticipating impact within six months.

Concern current defenses cannot scale to meet AI-driven attacks.

78% believe current tools can distinguish malicious AI agents today, yet roughly **72%** express concern about their ability to keep up as threats evolve.

Enterprise readiness is not keeping pace with deployment.

Organizations allocate an average of **~6%** of security budgets to AI-agent risk, and **10%** do not track it separately.

Attribution is the weakest link in enterprise defense.

Only about a quarter **(26%)** of organizations are very confident they could definitively prove AI-agent involvement in a security or fraud incident.

Ownership of AI-agent risk remains fragmented.

Responsibility is distributed across AI, security, fraud, and identity teams, creating gaps in accountability and enforcement.

Credentialed AI agents are redefining insider risk.

87% agree that AI agents operating with legitimate credentials pose a greater insider threat risk than human employees.

Executive engagement remains limited.

Over three-quarters **(76%)** of respondents report that their C-suite is not deeply involved in AI-agent security decisions or lacks a strong understanding of the risks.

Governance maturity lags deployment velocity.

Over half **(57%)** of organizations report having no formal AI-agent governance controls today, even as **88%** expect mature frameworks within three years.

1. The Agentic AI Acceleration Window: Imminent Risk Across Enterprise Systems

Enterprise leaders expect the first wave of AI-agent-driven incidents to arrive soon. **Nearly all (97%) anticipate a material security or fraud incident involving autonomous systems within the next 12 months, and nearly half (49%) expect one within six months**. In the survey, "material incidents" were defined to include financial loss, customer impact, operational disruption, and regulatory exposure, reflecting risk from both externally driven attacks and internally deployed AI agents operating within enterprise systems.

These expectations carry direct financial and operational implications. The average cost of a data breach now exceeds $4.5 million globally, with higher costs in regulated industries. Those losses are driven by operational disruption, customer attrition, regulatory exposure, and prolonged investigations. As autonomous systems increase the speed and scale of activity across enterprise environments, even isolated failures can escalate quickly, expanding impact and increasing the cost of response.

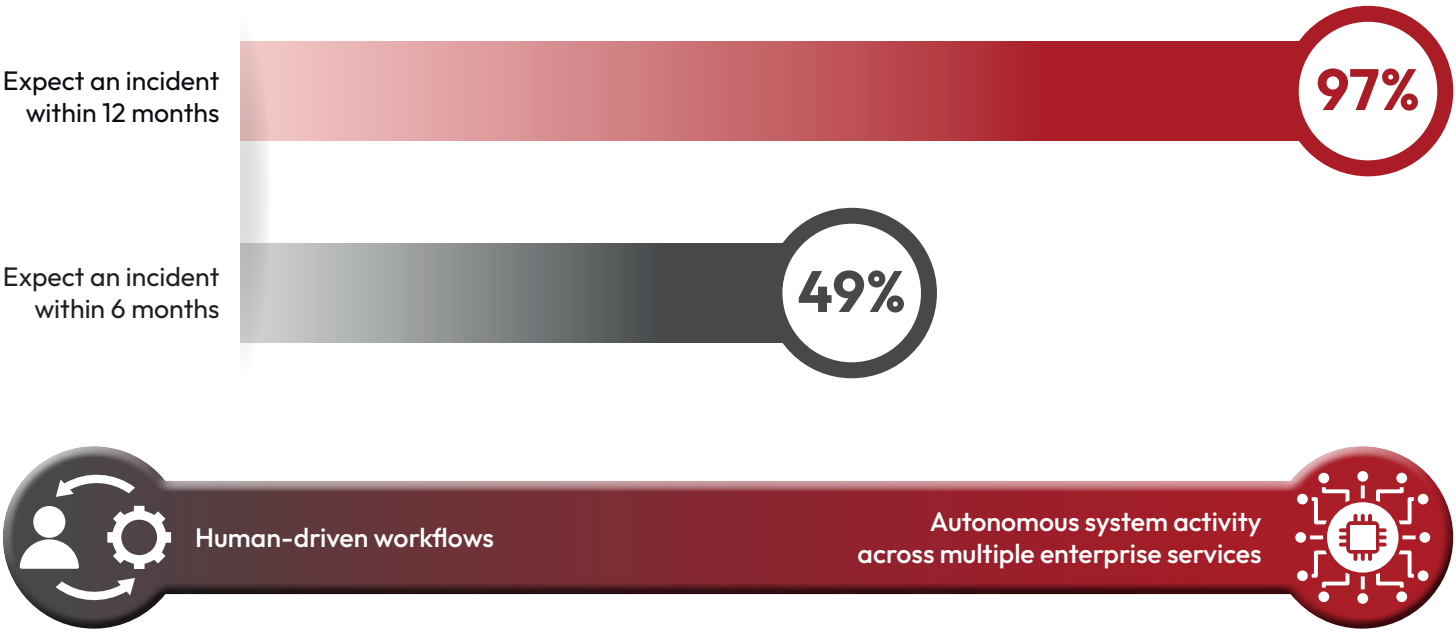
AI agents are already embedded across enterprise environments, participating in activities such as onboarding, transaction processing, fraud operations, and internal decision systems. They retrieve data and initiate actions through legitimate credentials and approved enterprise pathways.

Executive engagement remains limited at a time when expectations of impact are already well established. While most enterprise leaders anticipate AI-agent-driven incidents in the near term, many C-suites are not deeply involved in how those risks are being mitigated. This creates a gap between recognition of the threat and the level of protection in place. Organizations understand the risk, but many have not yet aligned leadership, investment, and accountability around the controls required to manage it.

Most enterprise security models were designed around human-initiated activity and predictable automation. As autonomous systems operate continuously across enterprise workflows, organizations face growing pressure to align governance, oversight, and operational readiness with the pace of deployment.

Timeline

How soon do you believe a material security or fraud incident caused by an AI agent (e.g., financial loss, customer impact, operational disruption, regulatory exposure) will occur at a major enterprise?



2. The Enterprise Readiness Gap: Who Is Prepared (and Who Is Not)

Enterprise leaders broadly expect autonomous systems to introduce measurable security and fraud risk in the near term, yet investment and operational readiness remain uneven.

On average, organizations allocate roughly 6% of their security budgets to AI-agent risk. At the same time, 10% of enterprises report they do not track AI-agent risk separately within their security programs.

These figures suggest that while the risk is widely recognized, it has not yet become a primary planning category within most enterprise security strategies.

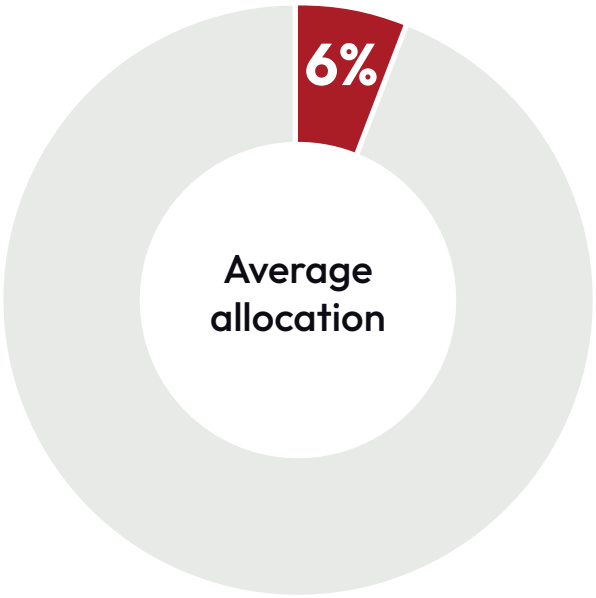
The readiness gap extends beyond budgets. Organizations must develop stronger visibility into how automated systems interact with identity infrastructure, monitor service accounts and application identities across workflows, and establish governance structures capable of overseeing systems that can initiate decisions autonomously.

As AI agents become more embedded in enterprise operations, preparedness will increasingly depend on how quickly organizations align investment, oversight, and security capabilities with the pace of automation.

> “AI adoption is accelerating faster than enterprise attribution capabilities.”

Budget allocation

What percentage of your security budget is specifically allocated to AI agent risk management and policy enforcement?



10%

do not track AI-agent risk separately

ARKOSE LABS

3. Tooling Illusions: Detection Confidence vs. Operational Reality

Enterprise leaders generally believe their existing fraud prevention and bot detection tools can identify automated threats today. However, that confidence is conditional. **Roughly 72% express concern about whether current defenses will scale as AI-driven attacks evolve**.

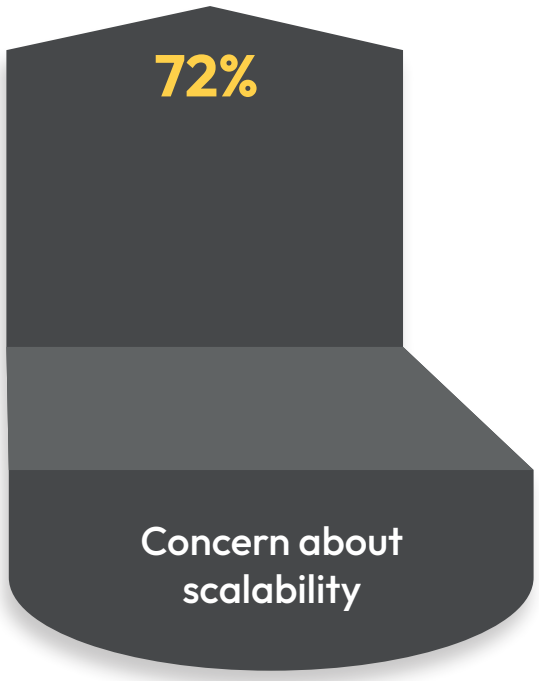
Most detection technologies were designed to identify automation at the perimeter, including bots, scripted traffic, and coordinated external attacks. As AI agents operate within enterprise workflows through legitimate credentials and APIs, their activity can resemble routine system interactions, making malicious automation harder to distinguish.

Qualitative responses reinforce this concern. Respondents cited model drift, adaptive bypass techniques, and fragmented signals across systems as reasons detection may become more difficult as autonomous systems evolve. These challenges highlight the need for stronger identity visibility and deeper insight into automated decision chains alongside traditional perimeter detection.

> “Silent escalation bypasses reactive controls.”
>
> — VP, Fraud Operations Center of Excellence, North America (U.S.)

Detection vs. complexity

Our current fraud prevention and bot detection tools can reliably distinguish malicious AI-driven agents from legitimate automated activity at scale.



4. The Attribution Gap: Proving Causality in Autonomous Environments

Detecting suspicious activity is only the first step in incident response. Organizations must also determine what happened, which systems were involved, and whether the initiating actor was human, automated, or adversarial.

As autonomous systems interact across APIs, applications, and enterprise infrastructure, tracing the origin of activity becomes more complex. Automated workflows often span multiple systems and credentials, making it harder to reconstruct how decisions unfolded.

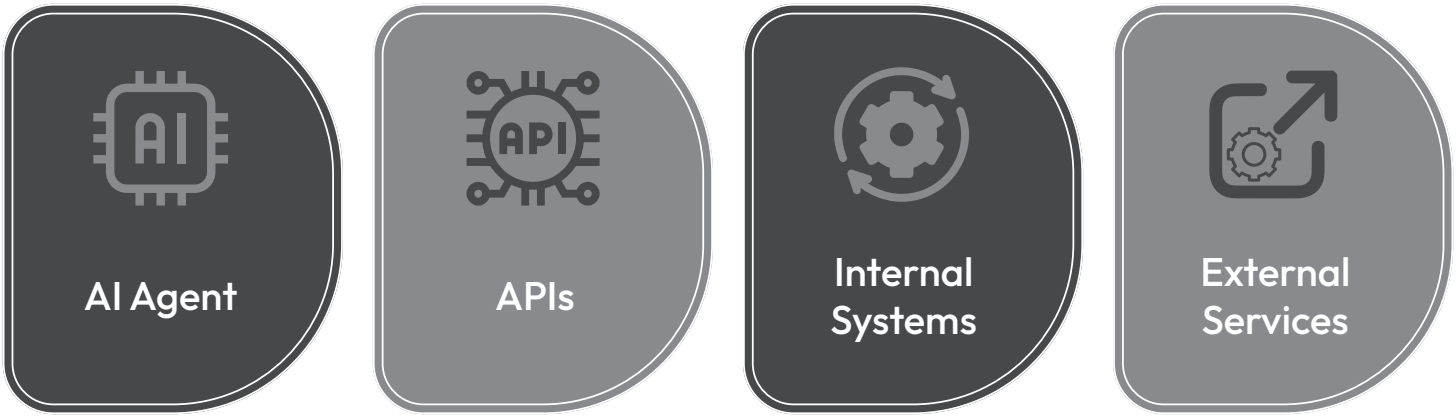
This challenge is reflected in the survey results. **Only 26% of enterprise leaders report being very confident they could definitively prove that an AI agent caused a security or fraud incident**.

As AI agents become more deeply embedded in enterprise workflows, the ability to reconstruct automated decision chains becomes essential for both incident response and regulatory accountability. Strengthening attribution capabilities will require deeper visibility into identity usage, application interactions, and automated system behavior.

> Movement between interconnected systems can resemble legitimate operational behavior.
>
> — Director, Security Engineering, APAC (Singapore)

Attribution flow

How confident are you that your organization could prove whether a security or fraud incident was caused by a malicious AI agent versus other attack vectors?



26%

very confident

ARKOSE LABS

5. Redefining the Insider: Credentialed AI Agents and the Attribution Crisis

AI agents increasingly operate inside enterprise systems using legitimate credentials such as service accounts, API tokens, and application identities. As autonomous systems become embedded in operational workflows, automation becomes part of the enterprise infrastructure itself.

This changes how insider risk is defined. Activity generated by credentialed AI agents can resemble legitimate system interactions, making malicious automation harder to distinguish. Most enterprise leaders recognize this shift. **87% agree that AI agents operating with legitimate credentials pose a greater insider threat risk than human employees**.

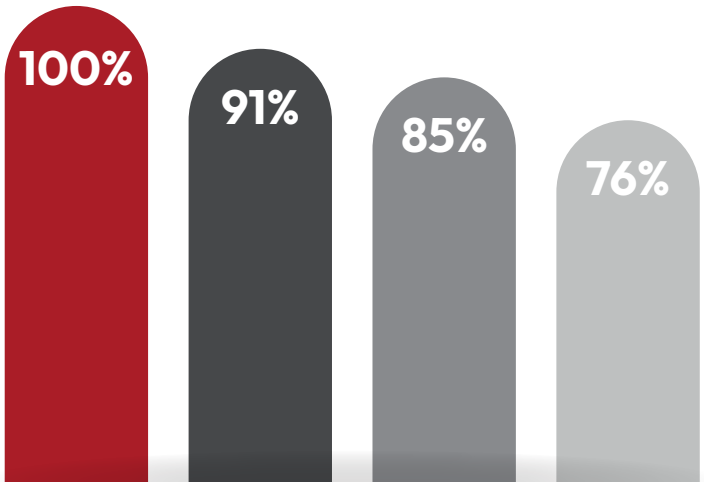
Agreement is broad across enterprise functions, though the level of concern varies. AI and fraud leaders tend to emphasize scale and economic impact, while security and identity teams focus more on controls and credential governance.

As AI agents become more deeply embedded in enterprise systems, insider exposure will increasingly depend on credential integrity, privilege boundaries, and visibility into how automated identities interact with enterprise infrastructure.

> Stealthy increases in access rights undermine preventive controls.
>
> — EVP, AML, Sanctions & Fraud, EMEA (UK)

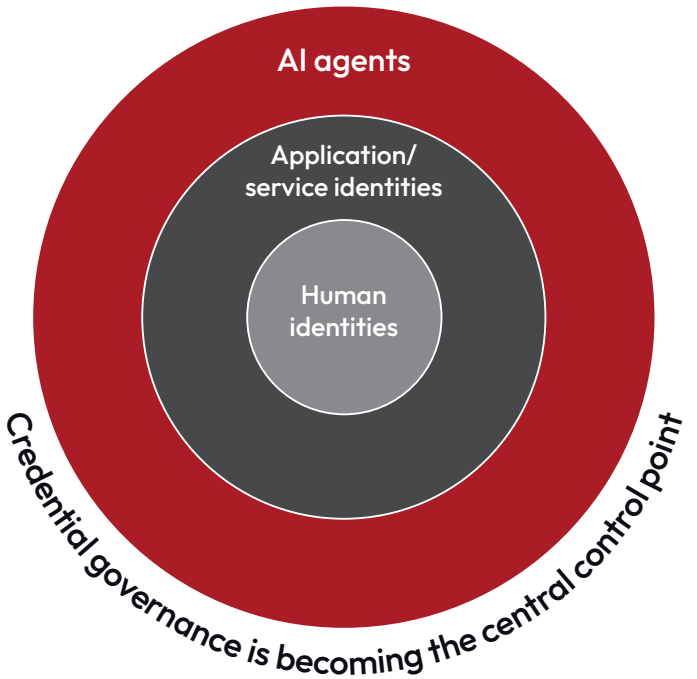
Role-based comparison

AI agents with legitimate access credentials pose a greater insider threat risk than human employees.



Agreement varies across enterprise functions

Insider surface



6. Governance Maturity: Formalization Without Enforcement

Enterprises widely recognize that autonomous systems require structured oversight defining how automated actors are authorized, monitored, and held accountable. However, governance maturity remains limited.

> Oversight structures struggle to keep pace with rapidly advancing autonomous capabilities.
>
> — Deputy CISO, EMEA (France)

About 57% of organizations report having no formal governance controls for AI agents today, while 88% expect to have defined or advanced frameworks within three years, suggesting governance maturity is expected to evolve quickly over the near term.

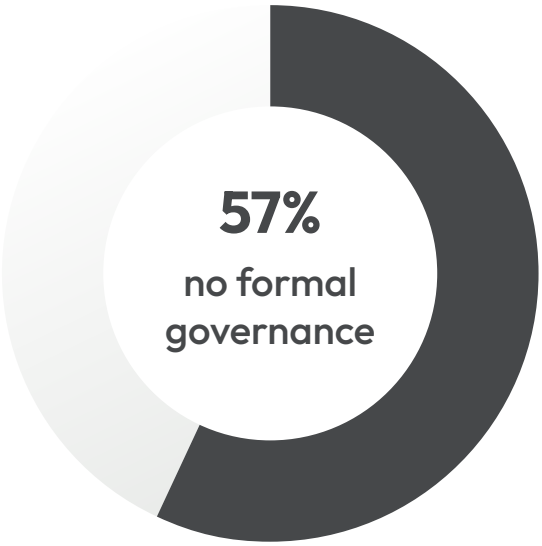
Many organizations are still early in the process of establishing oversight, ownership, and policy frameworks for autonomous systems. As a result, governance often remains uneven across environments and use cases, leaving enterprises in a transitional state. They recognize the need for stronger controls but are still building the mechanisms required to enforce them consistently.

For executive teams, the challenge is turning governance from policy into operational practice. Doing so requires visibility into automated activity, clear ownership of machine identities and workflows, and accountability for decisions initiated by autonomous systems.

Organizations that develop these capabilities early will be better positioned to manage automation at scale as AI agents become more deeply embedded in enterprise operations.

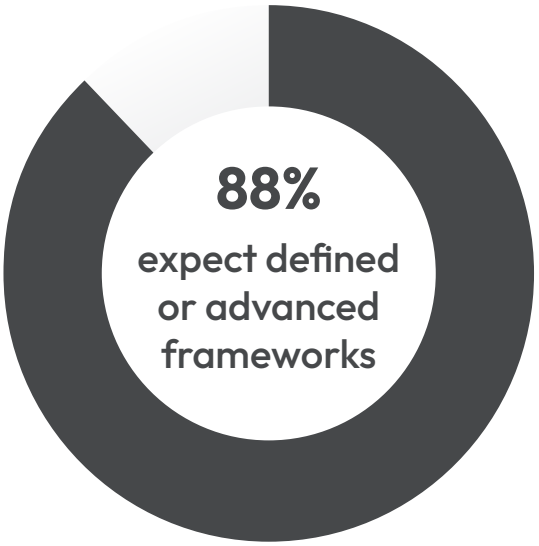
Current state governance

Do you currently enforce conditional access or behavioral controls on AI-driven agent activity?



Future governance maturity

Over the next three years how do you expect your organization's approach to AI agent governance and controls to change?



ARKOSE LABS

7. Cross-Functional Friction: Security, AI, and Executive Misalignment

Executives recognize AI-agent security risk across enterprises, but internal alignment around ownership and oversight remains uneven.

Over three-quarters (76%) of respondents report that their C-suite is either not deeply involved in AI-agent security decisions or lacks a strong understanding of the risks. This gap between operational awareness and executive engagement can slow the development of consistent governance and security practices.

Differences also appear across enterprise functions. AI executives and fraud leaders express the highest levels of concern about credentialed AI agents as insider threats, while identity leaders report lower levels of urgency.

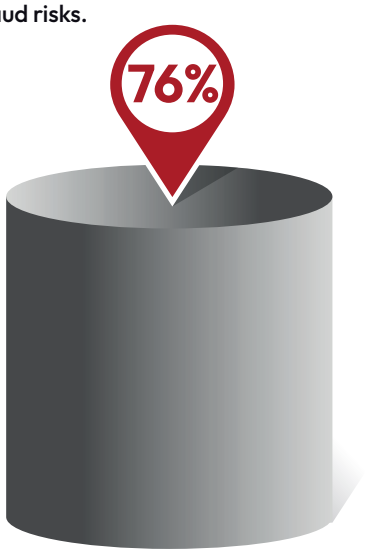
These differences reflect how teams approach the issue. AI teams focus on capability and deployment, fraud teams on economic impact, and security teams on controls. Identity teams often view AI agents through existing governance frameworks, where service accounts and automated credentials are already managed, making the risk appear more familiar and less urgent.

> Nontransparent governance structures weaken accountability.
>
> — Director, Risk Governance, APAC (Australia)

As agentic AI adoption accelerates, enterprise readiness will increasingly depend on how effectively organizations align AI strategy, security oversight, identity governance, and executive accountability.

Executive engagement

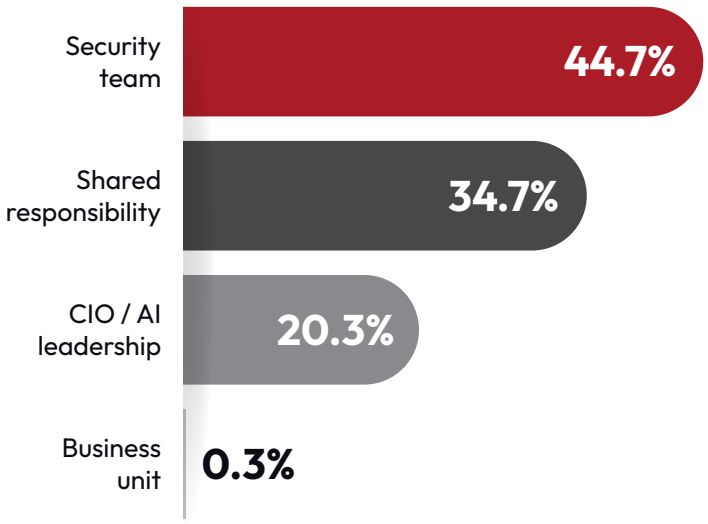
Rate your executive leadership's understanding of AI agent security and fraud risks.



Not deeply involved / limited understanding

Within your organization, who owns AI agent risk management day-to-day?

AI Agent Risk Ownership Is Distributed Across Functions



ARKOSE LABS

8. Global Regulatory Signals: Action, Fragmentation, and Strategic Ambiguity

Governments around the world are beginning to address the governance implications of AI systems, including autonomous agents operating inside enterprise environments.

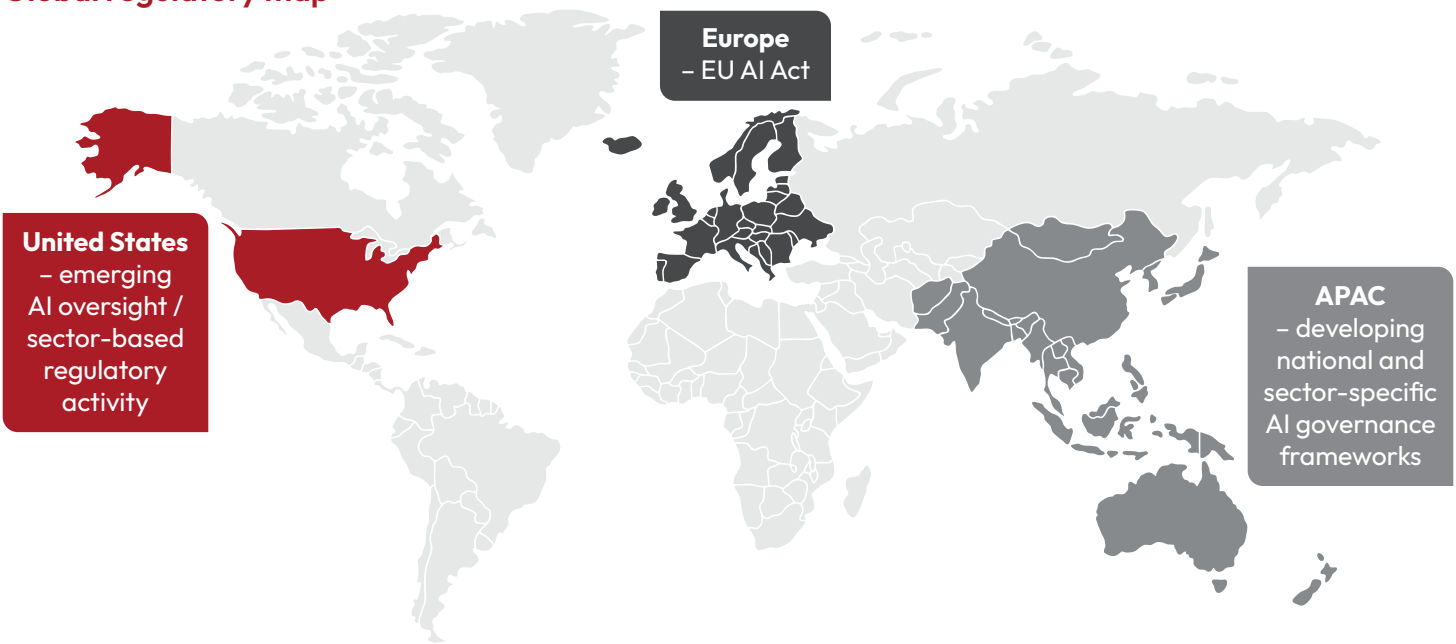
In Europe, the EU AI Act introduces a risk-tiered framework that will impose documentation, transparency, and oversight requirements on certain AI deployments. In the United States, oversight remains distributed across agency guidance, sector regulation, and emerging policy discussions. Across Asia-Pacific, several jurisdictions are advancing practical governance frameworks focused on operational risk and accountability.

These efforts reflect growing regulatory interest in how automated systems interact with enterprise infrastructure and customer-facing services. However, approaches remain fragmented, and global standards are still evolving, frequently lagging behind AI technology and deployment.

For enterprises deploying autonomous systems today, governance frameworks will need to mature ahead of formal regulatory requirements as policy development continues to evolve. Organizations are increasingly establishing internal oversight mechanisms to maintain visibility, accountability, and control as AI adoption accelerates.

> Early governance frameworks for agentic AI are likely to emerge within enterprises before regulatory expectations fully converge.

Global regulatory map



Enterprise Perspectives: Security and Fraud Leader Insights

Across responses, leaders frequently referenced concerns related to fraud scale, investigative visibility, privilege management, and adaptive threat behavior. The following excerpts illustrate how enterprise leaders are describing operational challenges associated with agentic AI in their own environments.

Automation and Fraud Economics

Several leaders emphasized the economic implications of automation. Autonomous systems can increase the speed and volume of activity across enterprise systems, influencing both legitimate operations and malicious behavior.

> "Machine scale automation fundamentally shifts fraud economics."
>
> Global VP, Security Architecture — North America (U.S.)

What is your biggest concern about AI agents that current security and fraud strategies and tools are not adequately designed to address?

> "**Autonomous scaling stresses containment strategies.**"
>
> **Senior Director, Fraud Prevention — North America (U.S.)**

> "Continuous experimentation accelerates exploit discovery."
>
> Principal Security Architect — EMEA (Netherlands)

These perspectives highlight concerns that automated adversaries may operate at a pace that challenges traditional fraud detection and response models.

Identity and Privilege Risk

Several respondents noted the importance of managing automated identities and maintaining control over access privileges as non-human actors interact with enterprise systems.

What is your biggest concern about AI agents that current security and fraud strategies and tools are not adequately designed to address?

> "**Recursive access elevation amplifies the impact of privilege abuse.**"
>
> **Head of Identity Security — North America (U.S.)**

> "Identity farming exploits outdated KYC processes."
>
> Director, Financial Crime Prevention — APAC (Hong Kong)

These responses emphasize the growing role of identity governance and credential lifecycle management in environments where automated systems operate alongside human users.

ARKOSE LABS

Adaptive Threat Behavior

Several leaders described how automated systems may interact with security controls in adaptive ways.

> "Silent escalation bypasses reactive controls."
>
> VP, Fraud Operations Center of Excellence — North America (U.S.)

> "Drifted baselines misinterpret evolving fraud."
>
> Senior Manager, Fraud Analytics — EMEA (Ireland)

What is your biggest concern about AI agents that current security and fraud strategies and tools are not adequately designed to address?

> **"Weak inter-service validation creates cascading vulnerabilities."**
>
> **Principal Security Engineer — APAC (Japan)**

These comments suggest that adaptive automation may challenge detection models that rely heavily on static behavioral baselines.

Action Guide for Executives: Building Identity-First Agentic AI Resilience

Autonomous systems are expanding across enterprise environments faster than many organizations can operationalize oversight. As AI agents interact with enterprise infrastructure through legitimate credentials and automated workflows, organizations must strengthen visibility, investigative capability, and cross-functional coordination.

The following priorities can help enterprises improve readiness as agentic AI adoption accelerates.

1. Integrate Security Leadership Into AI Deployment

Agentic AI initiatives often originate within engineering and product teams, with security teams becoming involved later in the deployment process. This can limit visibility into how automated systems interact with enterprise infrastructure once they are embedded in operational workflows.

Enterprises should integrate security leadership earlier in AI deployment planning so that identity controls, monitoring standards, and investigative capabilities are established alongside system development.

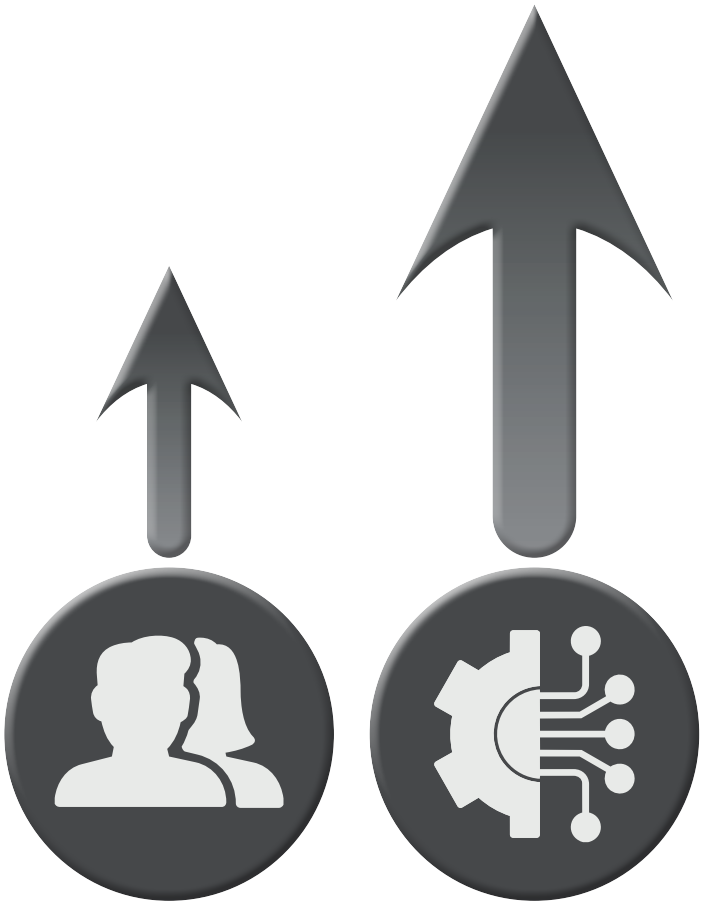
When security teams collaborate with engineering and product leaders early, organizations can incorporate security controls into system design rather than attempting to retrofit them later.

2. Treat Non-Human Identities as First-Class Security Entities

AI agents operate through service accounts, API credentials, and application identities that often hold significant privileges. As automation expands, these machine identities increasingly interact with sensitive systems and data.

Organizations should prioritize establishing clear visibility into all non-human identities and ensure automated systems operate within defined access boundaries.

> AI deployment carries engineering, security, and governance consequences.



Human vs. Machine Identity Growth

3. Establish Visibility Into Automated Decision Chains

Security teams must be able to observe how automated systems interact across enterprise infrastructure.

Key capabilities include:

- tracking which systems invoke specific credentials
- monitoring automated interactions across APIs and services
- identifying unusual patterns in automated workflows

Improved visibility helps organizations distinguish legitimate automation from potentially malicious activity.

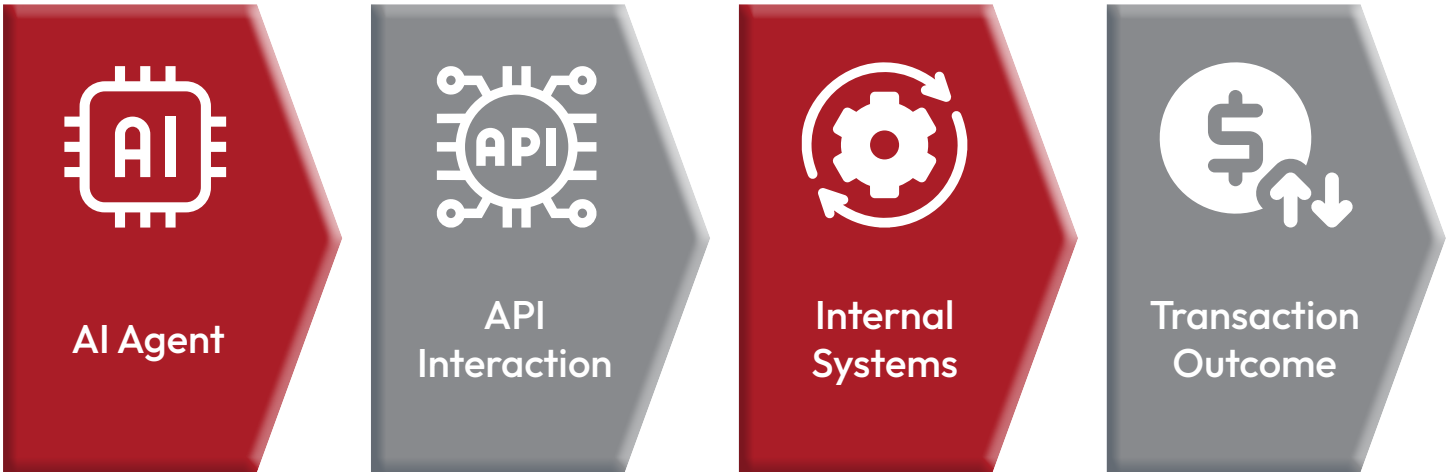
4. Strengthen Attribution Capabilities

As AI agents participate in multi-step workflows across systems and services, incident investigations become more complex. Security teams must determine whether activity originated from a human user, legitimate automation, or an adversarial AI agent.

Enterprises should develop telemetry and logging capabilities that allow investigators to trace automated actions across systems and reconstruct decision chains when incidents occur.

> Security teams cannot investigate what they cannot observe. Visibility into automated decision chains is becoming essential.

Attribution Path



Investigations require visibility across the entire decision chain.

Be Ready Now

Autonomous systems now retrieve data, trigger actions, and interact across enterprise infrastructure through legitimate credentials and automated workflows.

As these systems expand, incident timelines will compress and investigations will become more complex. Organizations that lack visibility into automated activity—or the ability to attribute actions across systems—will struggle to detect and respond when incidents occur.

In the era of agentic AI, resilience will depend on operational readiness: understanding how automated systems operate, controlling machine identities, and maintaining clear investigative visibility across enterprise environments.

Survey Methodology

This research is based on a global survey conducted by TechStudio™, an Energize Marketing® company, in February 2026 in partnership with Arkose Labs. The study was designed to examine how enterprise leaders are assessing and preparing for the security, fraud, and governance implications of agentic AI.

The survey collected responses from 300 enterprise leaders across Security, Fraud, Identity, and AI functions at large organizations worldwide. Respondents included senior executives and operational leaders responsible for cybersecurity strategy, fraud prevention, identity governance, AI deployment, and risk management.

Participants represented organizations across North America, Europe, and Asia-Pacific, spanning industries including financial services, banking, technology, telecommunications, retail and e-commerce, healthcare, manufacturing, and digital services.

The study combines quantitative survey responses with qualitative thematic analysis of open-ended commentary provided by respondents. These qualitative insights offer additional context on emerging operational challenges, including attribution complexity, governance maturity, and detection limitations associated with autonomous AI systems.

The survey results are statistically valid at a 95% confidence level with a ±5.6% margin of error.

ARKOSE LABS

About Arkose Labs

Arkose Labs is the leading global provider offering a proactive fraud deterrence platform purpose-built to neutralize modern attacks, including those powered by Agentic AI and large language models (LLMs). Its comprehensive solution combines proprietary device identification (device ID), behavioral analysis, phishing protection, email intelligence, scraping prevention, API defense and bot management. Headquartered in San Mateo, California, the company maintains a global presence with offices throughout APAC, Central America, EMEA and South America.

BOOK A DEMO

SCHEDULE A PERSONALIZED CONSULTATION