Gravitee

Adoption is outpacing control. — October 2026

The State of AI Agent Security 2026

AI agents are in production. Accountability is not. Our third survey of 900 technology leaders finds that 9 in 10 organizations have no one formally accountable for their agents, and most cannot trace or stop what those agents do.



EXECUTIVE SUMMARY

AI agents are in production. Accountability is not.

Accountability for an AI agent requires six things: it has its own identity, a named person answers for it, its purpose and limits are defined, its access is enforced, its actions can be traced, and it can be stopped when something goes wrong. Our October 2026 survey finds each of those missing for most organizations. And the gap is already being tested: 52% had an agent security or privacy incident in the past year, and only 1 in 4 of them then held a named person responsible.

- Confidence is starting to crack. In six months, the share who feel very prepared for an agent incident fell from 47% to 29%, and the share who believe regulation adequately covers agent risk fell from 39% to 26%. Leaders are starting to see the gap for themselves.

IDENTITY

17%

give every agent its own identity; the rest share accounts, borrow human credentials or use static keys

OWNERSHIP

9.5%

have a named individual formally accountable for agent behavior

PURPOSE AND LIMITS

29.5%

define the scope of what an agent may access before it goes live

ENFORCED ACCESS

51%

of agents, on average, have policy enforced on what they are permitted to do

TRACEABILITY

1 in 8

can fully reconstruct what an agent did 30 days ago and whose authority it used

STOP AND CONTAIN

6%

can revoke a misbehaving agent's access to all systems within an hour

About the research

We surveyed 900 senior technology leaders at organizations with 250 or more employees across the UK and US, in six sectors, 8 to 19 October 2026. Trends compare with our April 2026 wave.

WHAT CHANGED SINCE APRIL 2026

Six months on, confidence is cracking. Capability has not moved.

Between April and October 2026, every measure of self-assessed readiness fell, several by double digits. Over the same period, the share of agents actually monitored and secured rose by just 1.4 points.

Self-assessment fell

% of organizations, April 2026 (n=750) vs October 2026 (n=900)



Capability barely moved

Mean share of each organization's agents, and % of organizations



Six months of incidents appear to have replaced optimism with realism. What has not yet followed is the investment in controls.

Large agent estates have more than doubled since April. Nearly 1 in 4 agents acts without human review.

16%
run more than 150 AI agents, up from 7% in April 2026

7%
now run more than 200 AI agents

23%
of agents, on average, have full autonomy, rising to 27% in government

Organizations running more than 150 AI agents, by sector

% of respondents in each sector, October 2026. Bar scale runs to 25%.



How do you expect your AI agent deployment to change in the next 12 months?

% of all respondents, October 2026 (n=900)



Three in four plan to add agents in the next year, and almost none plan to cut back.

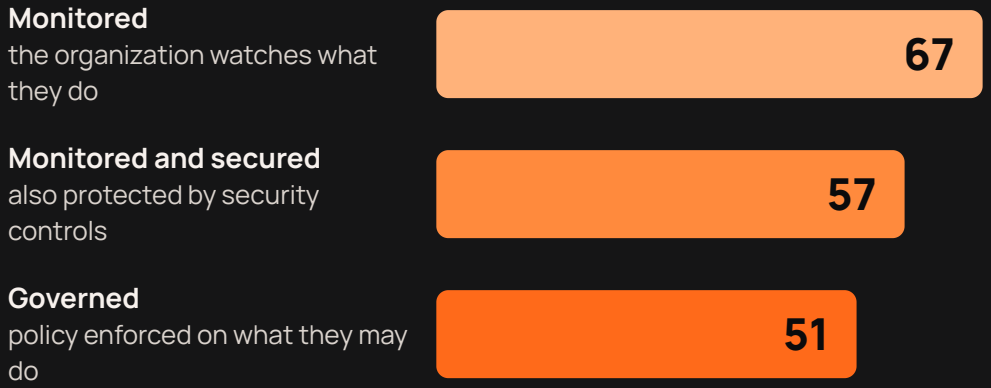
COVERAGE

Half of every agent fleet runs without enforced policy.

- New this wave, we separated three levels of coverage. Watching an agent is not the same as securing it, and neither is the same as enforcing what it is allowed to do. Each step loses ground: on average only 51% of an organization's agents have policy enforced on them.

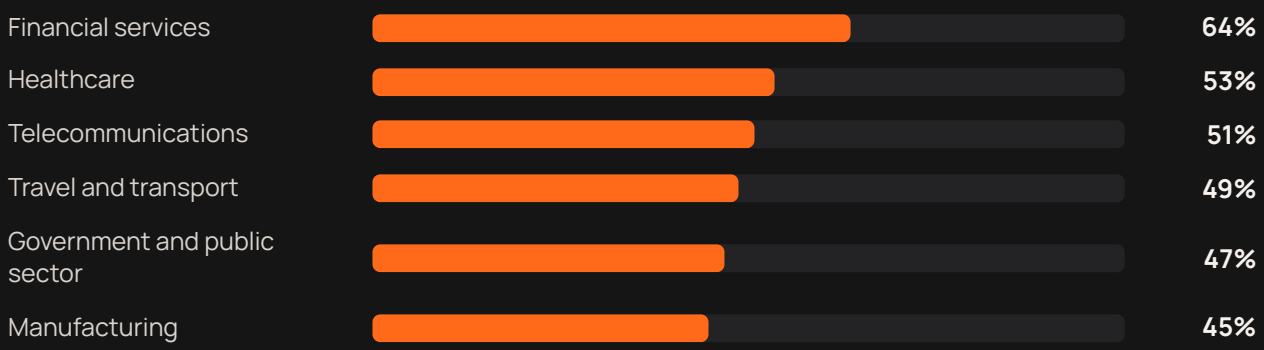
Of every 100 agents an organization runs, on average

Mean share of each organization's agents, October 2026 (n=893)



Share of agents governed, by sector

Mean share of each organization's agents with policy enforced, October 2026



Beyond the data

Visibility tells you what an agent did. Governance decides what it can do. **See how Gravitee enforces policy on every agent call.**

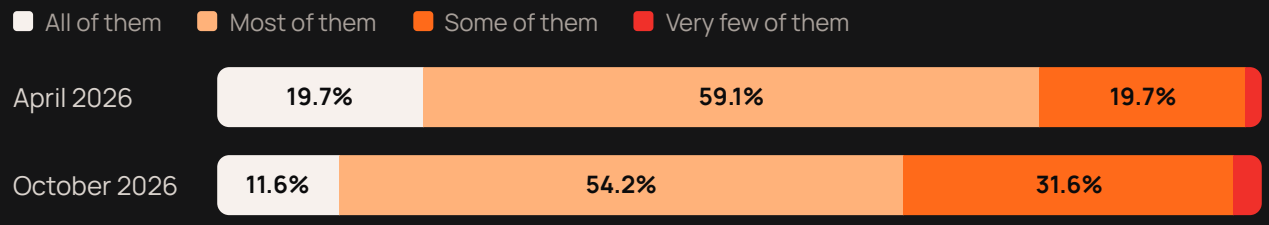
BEFORE GO-LIVE

Nearly 9 in 10 organizations put agents live before every one is secured and governed.

Only **11.6%** fully secure and govern every agent before it goes into production, down from 19.7% in April. The share admitting that only some or very few agents are covered rose from 21.3% to **34.2%**.

How many agents are fully secured and governed before going live?

% of organizations with agents deployed, April 2026 (n=748) vs October 2026 (n=893)



Very few of them: 1.6% in April, 2.6% in October. No organization said none.

Controls in place before an agent goes live

% of organizations with agents deployed, multiple answers allowed



No pre-launch control is used by more than about a third of organizations, and the ones tied to accountability lost the most ground: a named accountable person fell 6.9 points and a documented revocation process fell 3.9 points.

Only 17% give every agent its own identity.

Accountability starts with knowing which agent did what. Most organizations cannot tell, because their agents authenticate as something else: a shared service account, a static key, or the person who deployed them. In total, **45%** rely mainly on shared accounts, borrowed human credentials or static API keys, and when an agent uses a human's credentials, its actions are logged as that person's.

How AI agents generally authenticate to systems, data and APIs

% of organizations with agents deployed, October 2026 (n=893), one answer



Each agent has its own identity, by sector

% of organizations with agents deployed, October 2026



Beyond the data

Without a distinct identity per agent, delegation is opaque and audits lose meaning. **Explore the Agentic IAM Learning Hub** to see how to treat agents as first-class, identity-bearing principals.

97% say they are prepared for an agent incident. Only 6% can cut off a rogue agent within an hour.

97%
say they are prepared to respond to a serious agent incident

6%
could revoke an agent's access to all systems within an hour

46%
would need more than a day to revoke it

58%
have already revoked an agent's access during a real incident

How quickly could you revoke an AI agent's access to all systems and data?

% of organizations with agents deployed, October 2026 (n=893)



Within an hour 6.3% · Same day 47.1% · Within a few days 40.3%
Longer than a few days 6.0%, or no mechanism 0.2%

Could you reconstruct what an agent did 30 days ago, what data it accessed and whose authority it acted under?

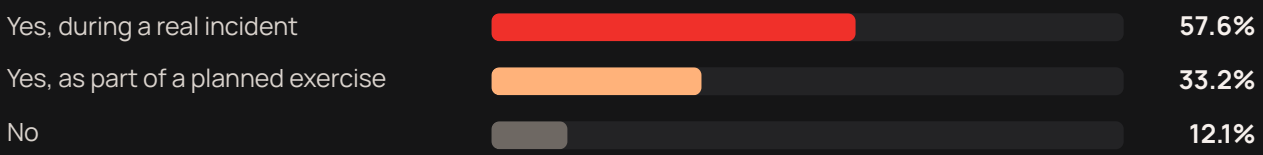
% of organizations with agents deployed, October 2026 (n=893)



Completely 12.5% · To a great extent 46.7% · To a moderate extent 33.0%
To some extent or very little 7.8%

Has your revocation process been used in the past 12 months?

% of organizations able to revoke an agent's access, October 2026 (n=891), multiple answers allowed



Revocation is not a theoretical control. More than half of organizations have already had to pull an agent's access in a live incident, yet for most it takes a day or more. One in eight can fully answer the basic audit question: what did this agent do, and on whose authority?

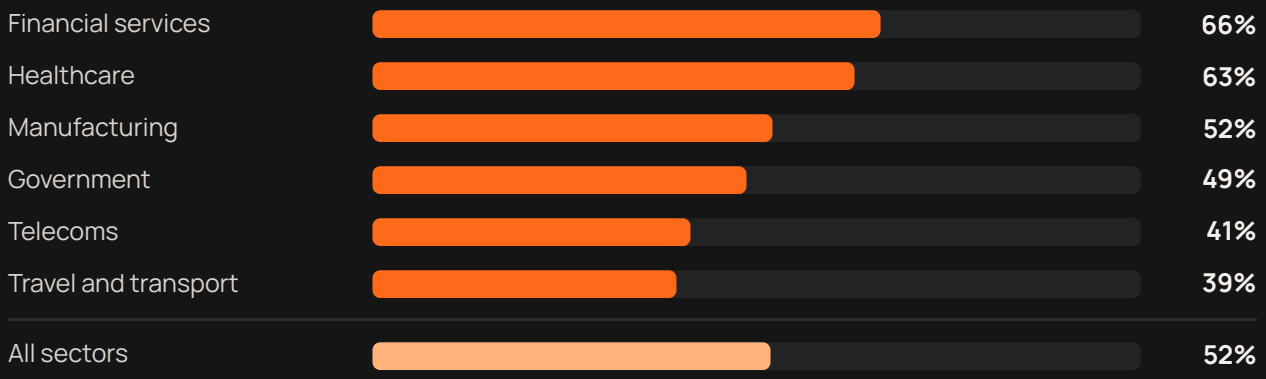
INCIDENTS

Half had an agent incident this year. Only 1 in 4 then held a named person responsible.

51.7% experienced or suspected an AI agent security incident in the past 12 months, and **21.4%** confirmed more than one. Financial services and healthcare lead. What happened next shows how thin accountability is when it is tested.

Incident in the past 12 months, by sector

Confirmed or suspected, % with agents deployed



What happened in the organization after the incident

% of organizations that had an incident, October 2026 (n=462), multiple answers allowed



After a real incident, the AI vendor was held responsible about as often as a named person inside the organization, and one in six organizations logged the incident and changed nothing.

IN THEIR OWN WORDS

Agent incidents are shifting: more agents are overstepping their access and sending data out.

Compared with April, more incident descriptions now involve agents reaching beyond their access or sending data outside the organization. Sensitive data exposed to the wrong people is still the most common pattern.

Agents acting outside their scope Up from 16% to 25% of incidents	*"It went out of its directive to perform a task."* *"Supply chain agent accessed unapproved files"*
Data sent outside the organization Up from 15% to 21%	*"agent leaked internal pricing data externally"* *"Agent leaked internal user credentials online"*
Sensitive data exposed to the wrong people Most common type of incident in both waves	*"Our customer service AI agent gave out incorrect product information and included a customer's email address in the response to another customer."* *"We suspected an AI booking assistant exposed limited customer travel details through an incorrect access permission"*
Regulated health data exposed Up from 2% to 13%	*"AI agents accessed sensitive patient data without authorization"* *"Agent shared restricted clinical data"*
Wrong or manipulated outputs acted on Steady at 17% in both waves	*"An internal customer service model generated inaccurate financial policy details during a live client interaction due to outdated training data."* *"AI agent followed malicious instructions."*

Beyond the data

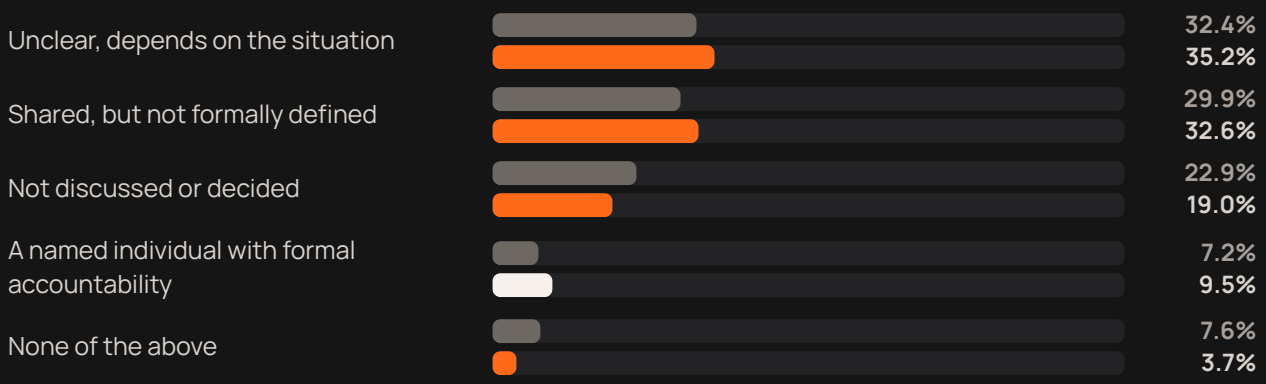
For a breakdown of the agentic risks behind these incidents, read our **practical review of the OWASP Top 10 for Agentic Applications**.

ACCOUNTABILITY

Fewer than 1 in 10 have a named owner for agent behavior. Ask who answers for harm, and the answers split five ways.

How is accountability for AI agent behavior handled?

% of organizations with agents deployed, April 2026 (n=748) vs October 2026 (n=893)



87%

have no formal accountability, up from 85% in April

CIO

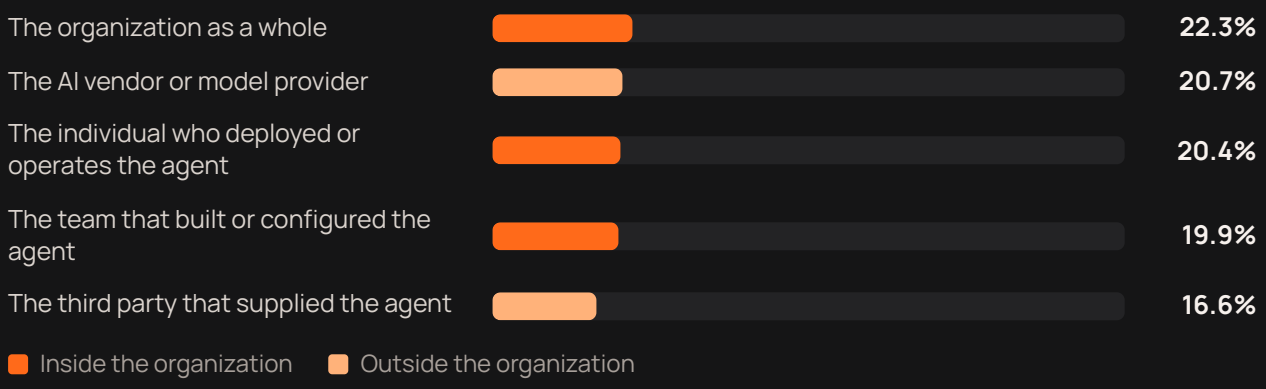
is the most common named owner where one exists (23.5%), ahead of the CTO and CISO

37%

say an outside party would be held most responsible: the AI vendor or the agent's supplier

If an agent caused material financial, legal or regulatory harm, who would be held most responsible?

% of organizations with agents deployed, October 2026 (n=893), one answer



AGENT-TO-AGENT CHAINS · NEW THIS WAVE

97% run agents that work through other agents. Only 27% define who is accountable across the chain.

97%
operate agents that invoke, delegate to or exchange data with other agents

40%
connect to agents that belong to other organizations (financial services: 55%)

16%
say accountability for these chains is simply unclear

Where one agent acts through another, who is accountable for the outcome?

% of organizations running agent-to-agent workflows, October 2026 (n=862)

28.3%
The initiating agent
Accountability sits with the agent that started the request

→

Intermediate agents
Delegated steps in between, where the work is actually done

→

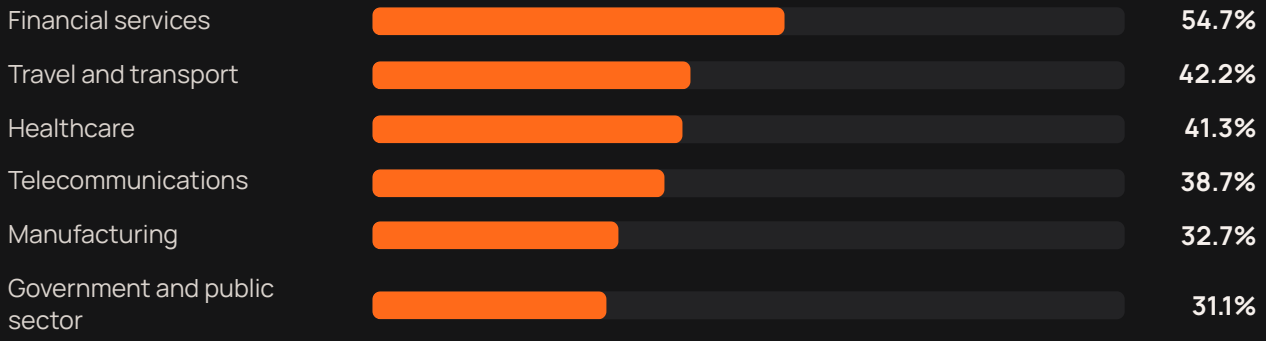
29.4%
The final agent
Accountability sits with whoever owns the agent that took the final action

26.5% define accountability formally across the full chain of agents. A further 15.8% say it is unclear.

Nearly three in four organizations with agent chains assign accountability to one end of the chain or cannot say where it sits, and they split almost evenly on which end. With 40% routing work through agents they do not own, that disagreement becomes a contract and liability problem, not only a technical one.

Agent-to-agent connections, by sector

% of organizations with agents deployed whose agents connect to other organizations' agents, October 2026

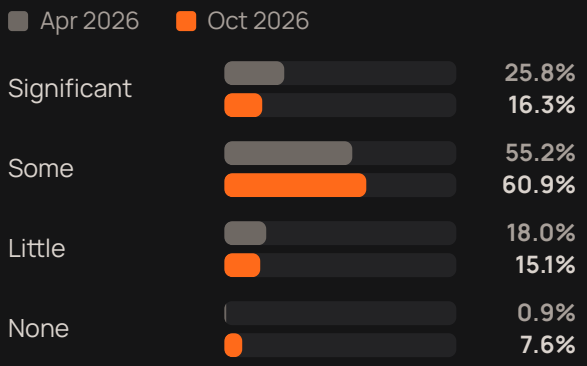


The pressure to ship has eased. The gap it created has not.

77% still feel pressure to deploy agents before security and governance are fully in place, but significant pressure fell from 25.8% to 16.3%. Productivity is the single biggest driver. Faith in regulation dropped further: only a quarter now think it covers agent risks.

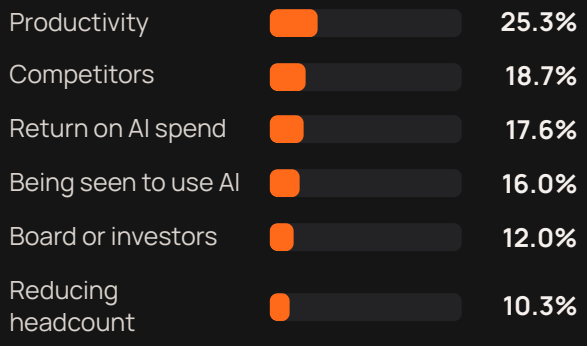
Pressure to deploy before governance is ready

% with agents deployed



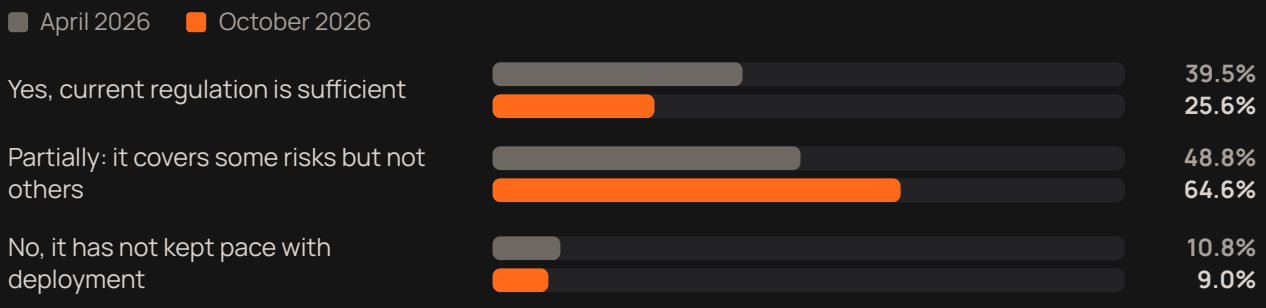
Single biggest driver of that pressure

% of those feeling pressure, Oct 2026 (n=825)



Does current national or international regulation adequately cover the risks of autonomous agents?

% of all respondents, April 2026 (n=750) vs October 2026 (n=900)



The share who think regulation is sufficient fell by 14 points in six months. The next page shows where leaders see the gaps, and what would raise their confidence.

Leaders are looking less to regulators and more to their own controls.

Asked what would most raise their confidence in deploying agents securely at scale, leaders put vendor transparency, internal skills and real-time visibility at the top. Clearer regulatory guidance fell from 23.9% to 19.9% and is now last. Among those who say regulation falls short, 27% point to accountability when an agent causes harm.

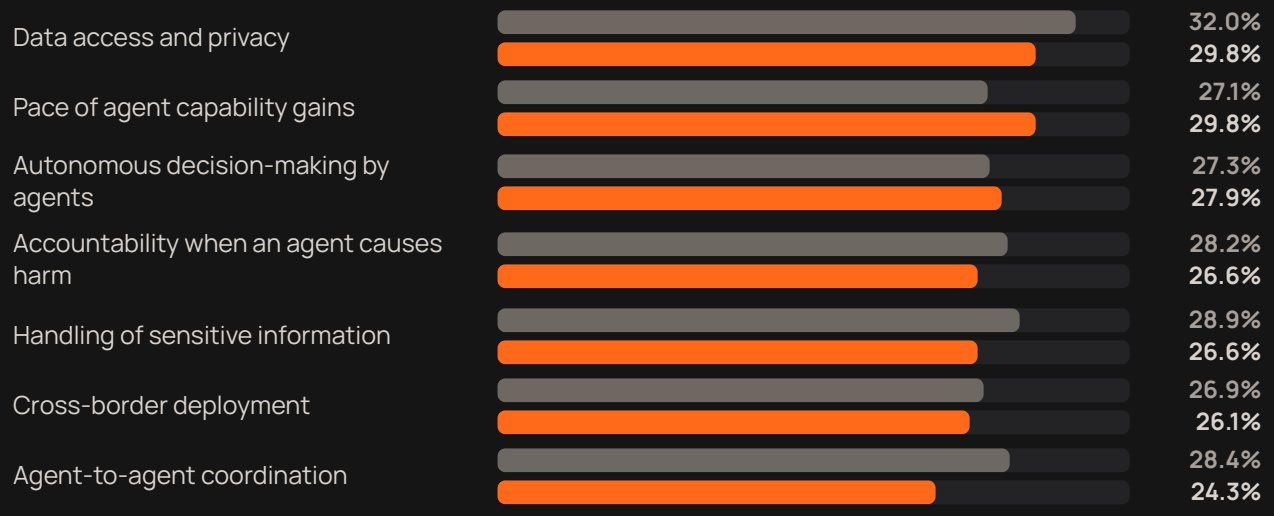
What would most increase your confidence in deploying AI agents securely at scale?

% of all respondents, Apr 2026 (n=750) vs Oct 2026 (n=900), up to two answers. Scale to 35%.



Where does regulation fall short?

% of those who say regulation falls short, Apr 2026 (n=447) vs Oct 2026 (n=662). Scale to 35%.



CONCLUSION

Accountability is the next control layer for AI agents.

Organizations understand the risk, and many have already lived it. An accountable agent needs six things in place, and today each one is missing for most organizations.

01	**Identity**	Every agent has a unique, verifiable identity, registered before it first acts.	**17%** give every agent its own identity
02	**Ownership**	A named person answers for each agent, and it is reassigned or retired when they leave.	**9.5%** have a named, formally accountable owner
03	**Purpose and limits**	Each agent has a documented purpose and explicit out-of-scope boundaries.	**29.5%** define an agent's access scope before go-live
04	**Enforced access**	Access is enumerated, derived from its purpose, bounded in time and revocable in real time.	**51%** of agents are governed, on average
05	**Traceability**	Every action is traceable to its sponsor, even across agent-to-agent chains.	**12.5%** can fully reconstruct a 30-day-old action
06	**Stop and contain**	Any agent can be stopped quickly, and what it did can be contained or rolled back.	**6%** can revoke access within an hour

READY TO CLOSE THE GAP?

Hold your agents to the Agent Accountability Framework.

An open framework of six testable principles, free for anyone to use. Ask its six questions of every agent you run, and help shape the standard.

Explore the Agent Accountability Framework

