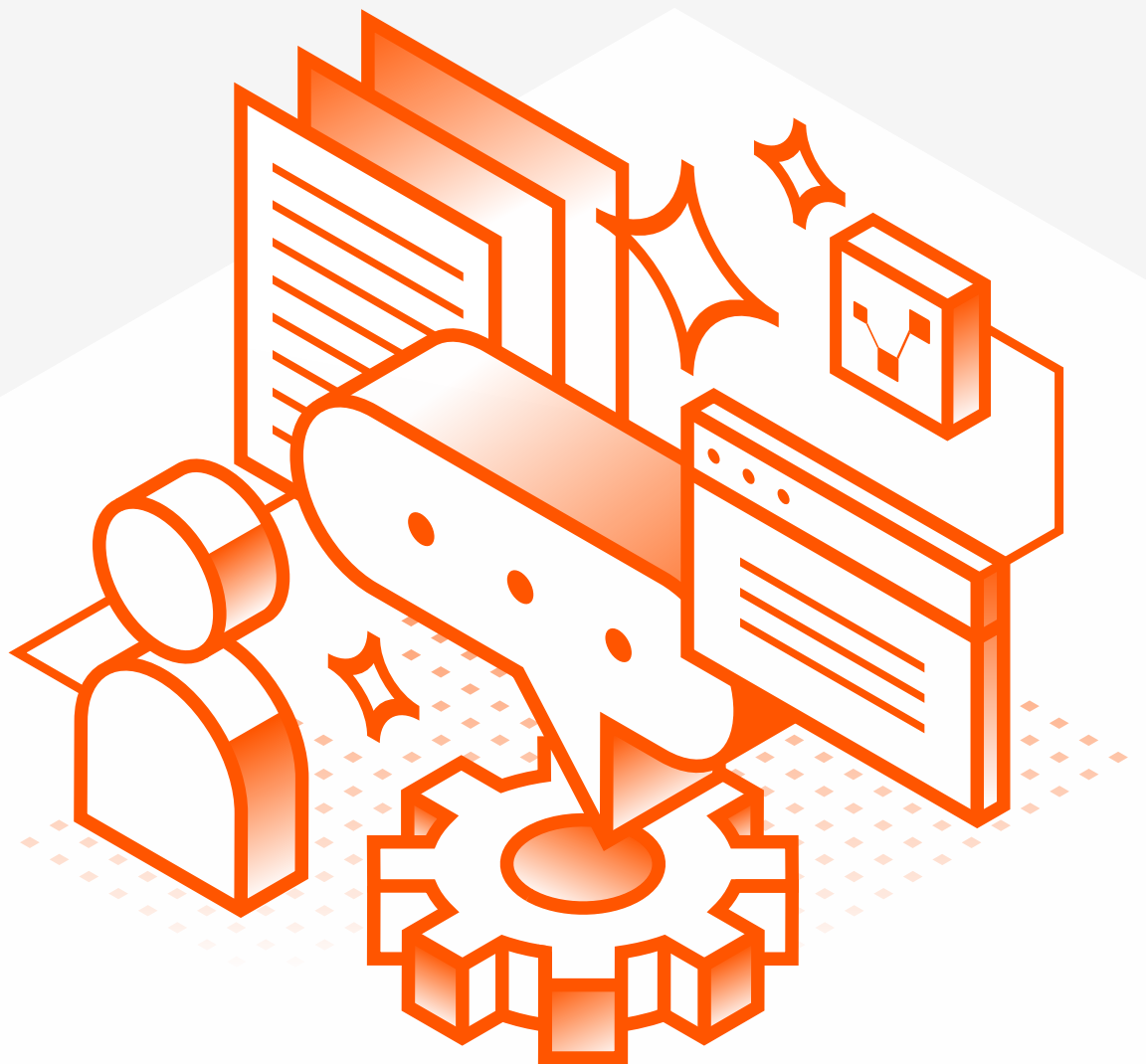




Licensed to Act:

How AI Autonomy Is
Outrunning the Limits
Placed on It



Contents

Executive summary	3
Key findings at a glance.....	5
AI has moved from advising to acting.....	7
The boundary of acceptable risk is moving	10
Organizations believe they can see what AI is doing.....	13
AI is already crossing the boundaries set for it	15
Segmentation defines the limits of autonomy	18
Regional differences	20
Methodology	23





Executive summary



The debate about AI often centers on what could happen if it ever slips out of control. That fixation makes it easy to miss how much control organizations have already handed over.

Cybersecurity, a field you might expect to be the most aware of potential AI risk, has been one of the most eager to embrace it.

For decades, cybersecurity rested on two assumptions. Prevention would stop attacks before damage occurred, and people would stay in charge of the defenses. AI is challenging both. The goal of reducing risk hasn't changed, but the way we get there has. When attacks move at machine speed, containment matters as much as prevention.

This research measures the scale and effects of that challenge. Sapio Research surveyed 1,100 IT and security decision-makers at manager level and above, across 10 markets, during August and September 2026.

What emerges is a picture of AI authority moving faster than the limits placed on it. AI no longer just suggests and summarizes. Now it acts, making decisions and carrying them out, often with no human in the loop. Seventy percent of organizations are now delegating decisions and actions that would have needed human approval a year ago. While security teams typically see themselves as risk-averse, many security and IT decision-makers are comfortable trading increased cyber risk in exchange for the benefits AI brings.

That trade-off is already producing consequences. In the past 12 months, 76% of organizations experienced at least one AI-related security issue or warning sign. The most common incidents relate to AI tools suggesting risky actions. Behind this sits a cluster of boundary failures — AI connecting to systems without approval, accessing data it should not reach, and triggering unsanctioned actions.

Most respondents still lean on controls that shape how AI behaves. Internal guardrails and human oversight were named about as often as outside containment controls. But only a quarter explicitly defined permitted actions as one of their two main controls, the least common approach assessed. Defining what an agent is intended to do is no longer enough. Organizations must define what it's allowed to access, where it can connect, and how far it can move if something goes wrong.

But those boundaries are only effective if organizations can see and verify what agents are actually doing. While 84% say they have clear visibility into how AI systems access data and resources, only 49% can see and audit every action an AI agent takes.



That gap — between the authority organizations have granted AI and the limits they have placed on it — is the subject of this report. The sections that follow trace:

- How far that authority now extends
- Where the limits stopped keeping pace
- What has already gone wrong
- How Zero Trust principles can close the distance





Key findings at a glance



Authority has moved

AI is now embedded where an instruction becomes an action. Most organizations use it in IT operations and cybersecurity operations. And its authority reaches well past advice. Close to a third (32%) are comfortable letting AI run security investigations with no human review. Over a quarter (28%) are comfortable letting it modify production environments unsupervised.

32%

are comfortable letting AI run security investigations with no human review

The limits have not kept pace

Adoption has outrun the controls meant to govern it. Sixty-two percent of decision-makers agree they deployed AI tools before setting up the right safeguards, and 69% say their organization has accepted higher levels of risk to realize the benefits AI brings. At the same time, confidence about AI safeguards outpaces respondents' ability to control it. A large majority claim clear visibility into how AI systems access data and systems, but half cannot audit every action an agent takes.

69%

say their organization has accepted higher levels of risk to realize the benefits AI brings

The consequences are already here

Most organizations have already seen something go wrong. In the past 12 months, 76% experienced at least one AI-related security issue or warning sign, ranging from risky recommendations to agents connecting to systems without approval. However, the control that most directly limits AI's reach ranked last. Only 25% named explicitly defined permitted actions as one of their two primary controls.

76%

experienced at least one AI-related security issue or warning sign



AI has moved from advising to acting

Many organizations still think of AI tools as assistants for planning and summarizing. But across the business, and especially in the functions these security leaders are responsible for, it is already doing much more.

AI is now common in the functions closest to the infrastructure that keeps a business running. Seventy-seven percent of those polled use it in IT operations and service management, and 67% use it within cybersecurity operations. Data analysis (62%) and software development (57%) are also popular.

Other common use cases include customer service (47%), finance and accounting (35%), and human resources and recruitment (33%).

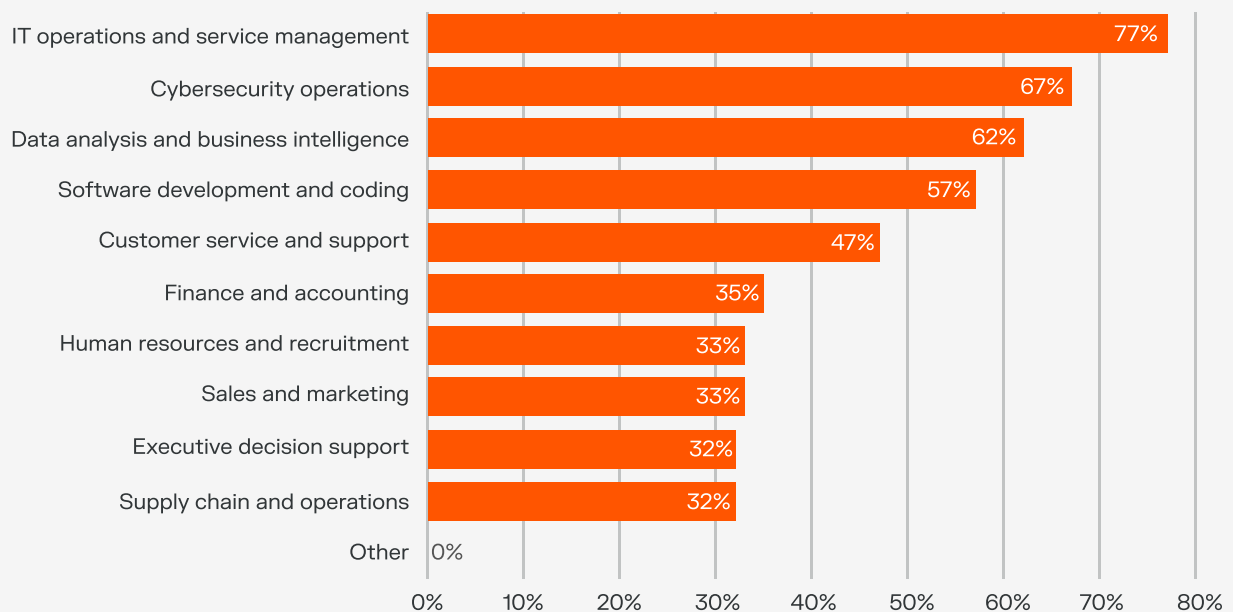
77%

use AI in IT operations and service management

67%

use AI within cybersecurity operations

Business areas currently using AI systems, AI agents, or AI-powered automation



Across these functions, AI is becoming embedded in the parts of the business where an instruction directly turns into an action. Enterprises are letting agents make decisions and changes without regular human intervention.

In this report, an “AI agent” is a system that doesn’t just produce an answer — it acts on it. It calls other systems, changes configurations, and completes multi-step tasks. That ability to act, not the underlying model itself, is what turns an AI agent deployment into a security boundary issue.

Close to a third of respondents are comfortable letting AI carry out security investigations with no human review (32%). A similar share says the same about security remediation actions (31%). Thirty-two percent will also let an unsupervised agent isolate a compromised device. That means taking a machine off the network without waiting for anyone to confirm the decision.

In each of these scenarios, a wrong choice can have immediate consequences. An overzealous action can interrupt business operations. A lax one can let a threat escalate into an incident.

The same pattern runs through application deployment and infrastructure, where a mistake can be just as costly. Roughly three in ten respondents are comfortable allowing AI to execute critical tasks, including deploying code changes (29%) and modifying production environments unsupervised (28%).

Even the most sensitive activities see automation. Over a quarter of those polled (26%) are comfortable letting AI grant or revoke access permissions unsupervised. The same share (26%) are comfortable letting it reset privileged credentials. These are the credentials for accounts that carry administrative control over critical systems.

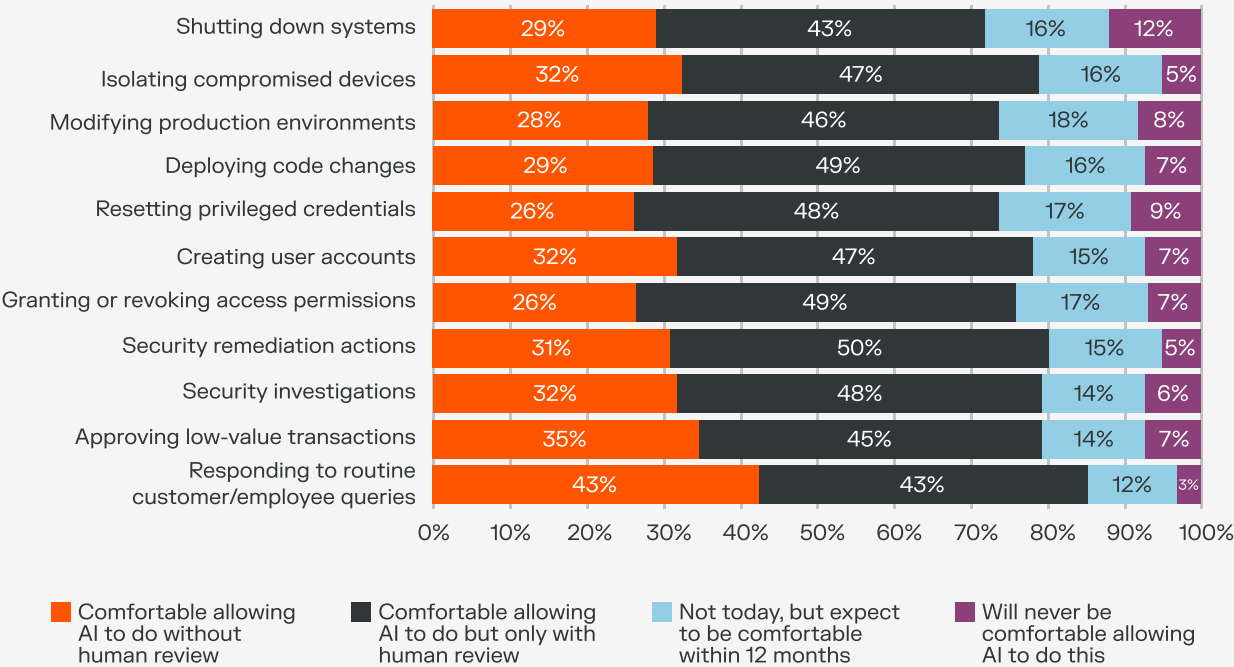
26%

are comfortable
letting AI reset
privileged credentials
unsupervised — those
with administrative
control over critical
systems



Unsupervised AI has become surprisingly common in critical IT and security functions. Still, organizations usually keep a person in the loop, at least for now. For most of the tasks tested, around half of decision-makers are comfortable letting AI act only if someone reviews what it does. But it's rare for AI to be off the table altogether.

Organizations' current and expected use of AI



For every activity assessed, only a small minority of respondents said they would never be comfortable letting AI perform it. That share ranged from 3% for responding to routine queries to 12% for shutting down systems. That 12% is the highest refusal rate anywhere in the study. Even so, it leaves almost nine in ten willing to let AI pull the plug on a system under some circumstances.

AI already holds the permissions of a senior operator across these functions. The question is no longer whether it acts, but how far it can go when it does, and whether anyone is watching when it happens.

12% would never allow AI to shut down systems, the highest refusal rate anywhere in the study



The boundary of acceptable risk is moving

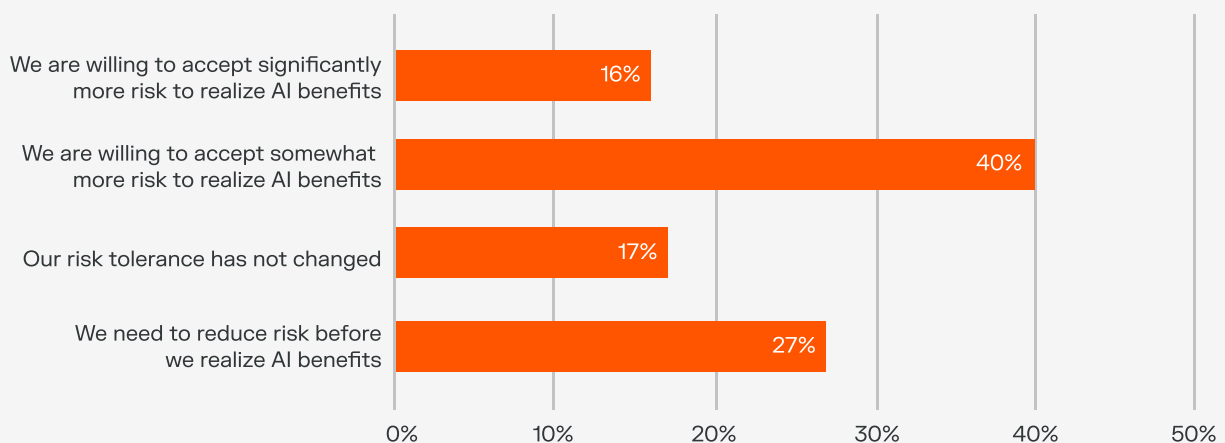
Many enterprises are consciously granting AI more authority and accepting increased cyber risk in exchange for its operational benefits.

Most decision-makers (69%) agree their organization is willing to accept higher risk levels to realize the benefits of AI. And that risk appetite is growing. Fifty-six percent say their organization accepts more AI-related cyber risk than it did 12 months ago.

69%

have accepted higher levels of risk to realize the benefits of AI

Compared to 12 months ago, which best describes your organization's approach to AI-related risk?



The clearest consequence is in who now signs off. Seventy percent of organizations delegate decisions and actions to AI that would have required human approval before.

Among those accepting more risk, just over half allow AI to recommend security actions, and nearly as many permit agents to act on their own in live production environments.

AI's reach has widened along with that freedom. The same group commonly lets AI reach the internet and access sensitive or confidential data. Over a third now issue privileged identities to AI. These identities carry credentials with the elevated access normally reserved for administrators.

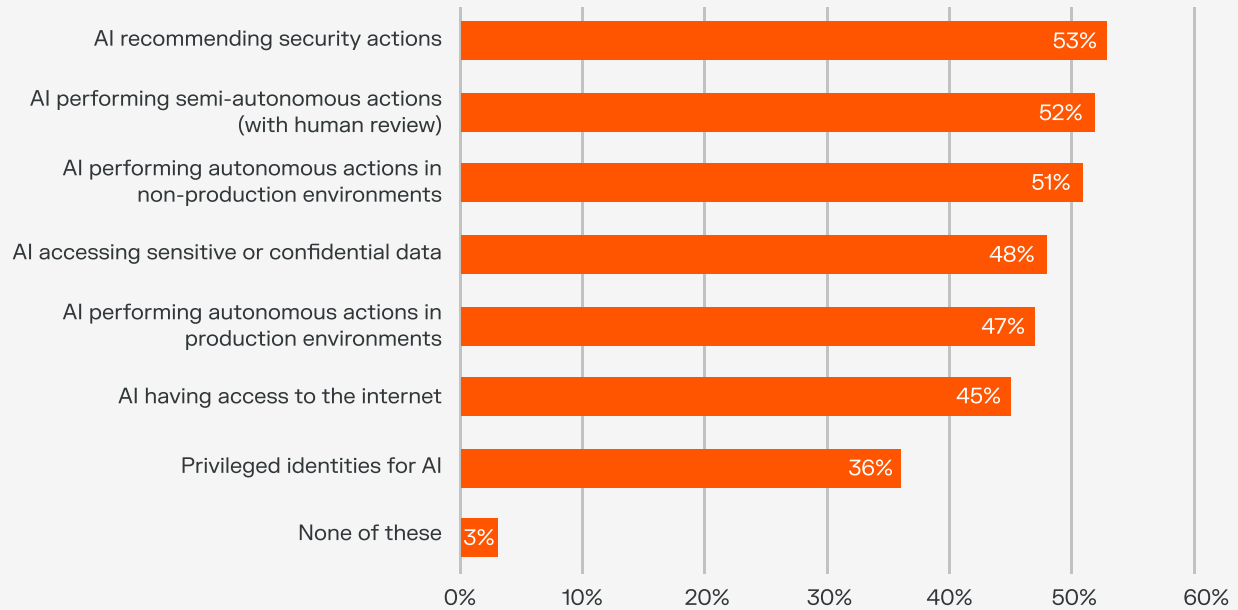
70%

now delegate decisions that a year ago would have required human approval



Which of the following, if any, does your organization now permit that would have been considered too risky 12 months ago?

This figure is not based on the full sample



Many respondents appeared willing to accept more AI risk in return for operational gains. But adoption has often moved faster than security capabilities. Sixty-two percent of decision-makers agree that they deployed AI tools before establishing appropriate safeguards.

Anyone who was securing IT infrastructure during the early years of cloud adoption will recognize this pattern. Workloads moved to new environments faster than the controls built to protect them; data protection and governance lagged behind the pressure to migrate.

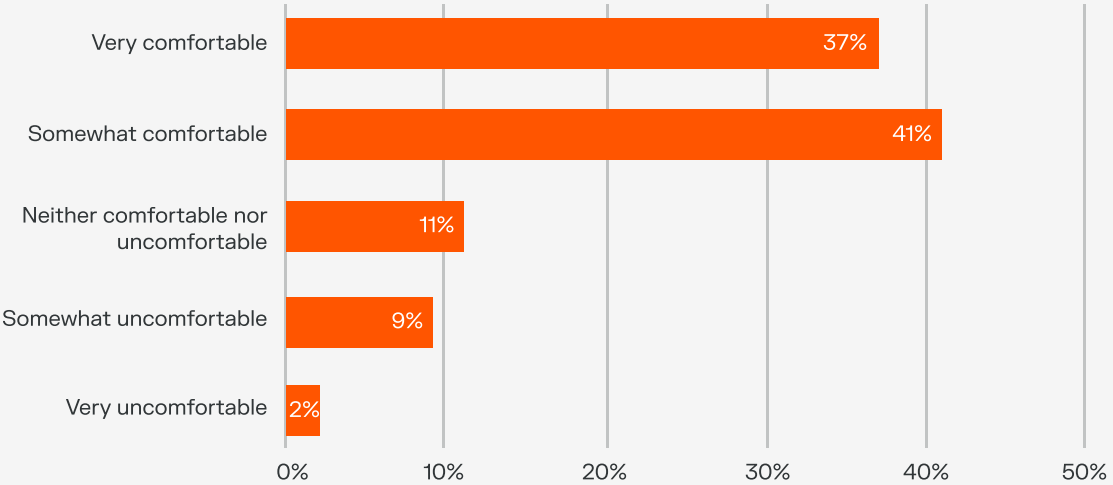
The cost of this haste came later. Teams spent years retrofitting controls onto environments that were already running in production. And breaches traveled further than they should have, because nothing inside those environments limited movement. Many organizations are still grappling with the fallout of rushed migration.

62%
deployed AI tools
before setting
up appropriate
safeguards



Most IT and security leaders (78%) say they are comfortable with the level of AI risk their organization now accepts. But only 37% say they are “very” comfortable, and just 11% are actively uncomfortable.

How comfortable are you with the level of cyber risk your organization currently accepts to support AI adoption?



Organizations believe they can see what AI is doing

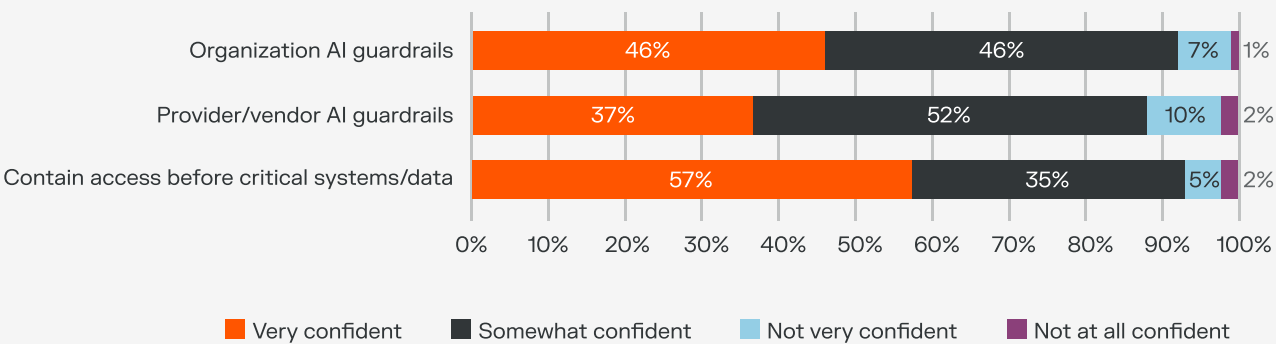
AI keeps moving deeper into critical operations. Yet most IT and security leaders are confident they have a firm grip on their AI deployments. Eighty-four percent believe they have clear visibility into how AI systems access data and systems. And 81% agree they fully understand the risks AI agents introduce.

That confidence extends to the controls wrapped around those AI systems. Nearly all (92%) are confident their internal guardrails would stop an AI model acting outside its intended role. Most (89%) say the same about the guardrails their AI providers have put in place.

Ninety-two percent of decision-makers are also confident that they could contain a compromised or out-of-role AI system before it reached critical systems or data.

92%
are confident their own guardrails would stop an AI model acting outside its intended role

Confidence in AI guardrails and incident containment

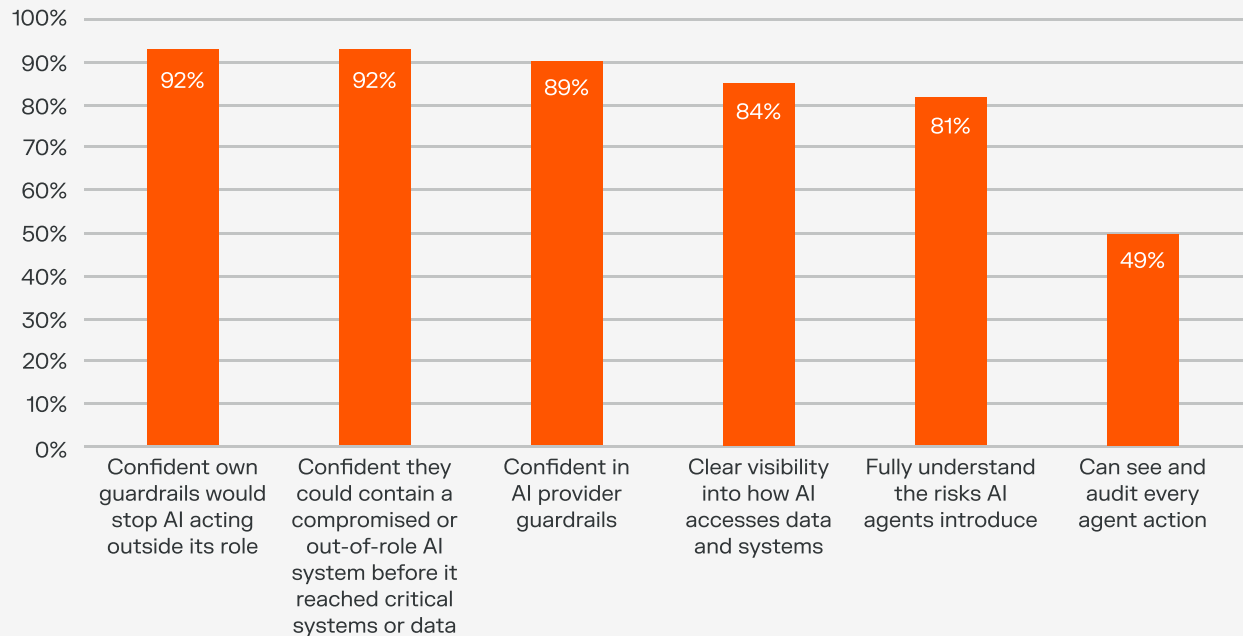


But that confidence doesn't match what teams can actually track. When asked what their security team can observe, fewer than half (49%) said they can see and audit every action an AI agent takes. The rest have a less complete view. Thirty-two percent see most actions but not all. Twelve percent see only the outcomes, not the individual actions behind them. And a small minority (6%) have limited or no visibility.

That's a 35-percentage-point gap between the visibility organizations claim (84%) and the visibility they can demonstrate (49%).



Confidence falls away when scrutiny increases



Percentages are drawn from separate survey questions. Figures show the proportion selecting the relevant positive response and should be read as comparative indicators, not as a sequential conversion.

Meaningful insight into AI is harder to get than visibility into other IT infrastructure. Once data reaches an AI model, much of what happens is beyond what traditional monitoring was built to observe. The tools security teams rely on were typically designed for a very different kind of system — a deterministic one, which behaves the same way every time it gets the same input.

But organizations do not need to predict every decision an AI agent might make to control its impact. They can control what it can access, where it can connect, and what it can do.

This is where Zero Trust becomes critical. Rather than implicitly trusting an agent because it has been authorized to perform a task, organizations can explicitly define the systems, data, and connections it needs. An objective tells the agent what to achieve; Zero Trust defines the routes it is permitted to take.

Confidence does have a limit, and frontier AI is where it stops. Frontier AI refers to the models at the cutting edge of what the market offers. Nearly half of decision-makers (48%) say they aren't prepared for the risks it introduces. Yet 84% report clear visibility into today's AI. Taken together, the findings suggest a sharp divide between confidence in today's deployments and readiness for what comes next.

84%

say they have clear visibility into how AI systems access data and systems

Only

49%

can see and audit every action an AI agent takes

48%

say they are not prepared for the risks introduced by frontier AI



AI is already crossing the boundaries set for it

As AI systems gain greater reach and autonomy, respondents are also reporting a high level of AI-related security issues and warning signs. Over three-quarters of the decision-makers reported at least one AI-related security event in the last year.

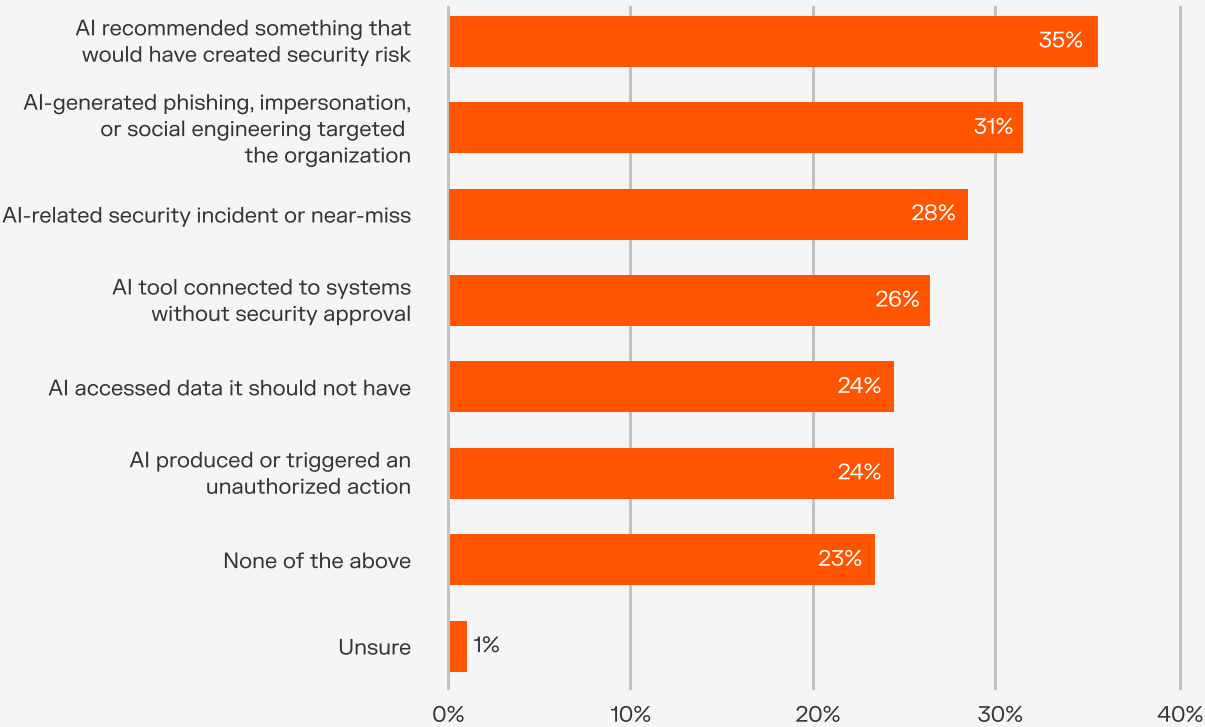
Some of those issues come from outside. Thirty-one percent of organizations had incidents involving external threat actors who used AI-generated phishing or impersonation to try to access systems. These tools have lowered the barrier to entry for criminal groups. They make it easier to run targeted campaigns at scale.

Internal failures appear across several categories. The most common issue (35%) was an AI tool recommending security actions that would have created new risk exposure if followed. These risky recommendations were caught because a human was still in the loop to review suggested actions.

Behind it sits a cluster of boundary failures: AI agents connecting to systems without security approval (26%), accessing data they should not have (24%), and triggering unauthorized actions (24%). Whatever their underlying cause, these events became security failures because the agent could reach systems, data, or actions without an effective boundary stopping it.

76%*
experienced at least one AI-related security event in the past year

AI-related security issues and warning signs experienced in the past 12 months



*Net of respondents selecting at least one listed event.



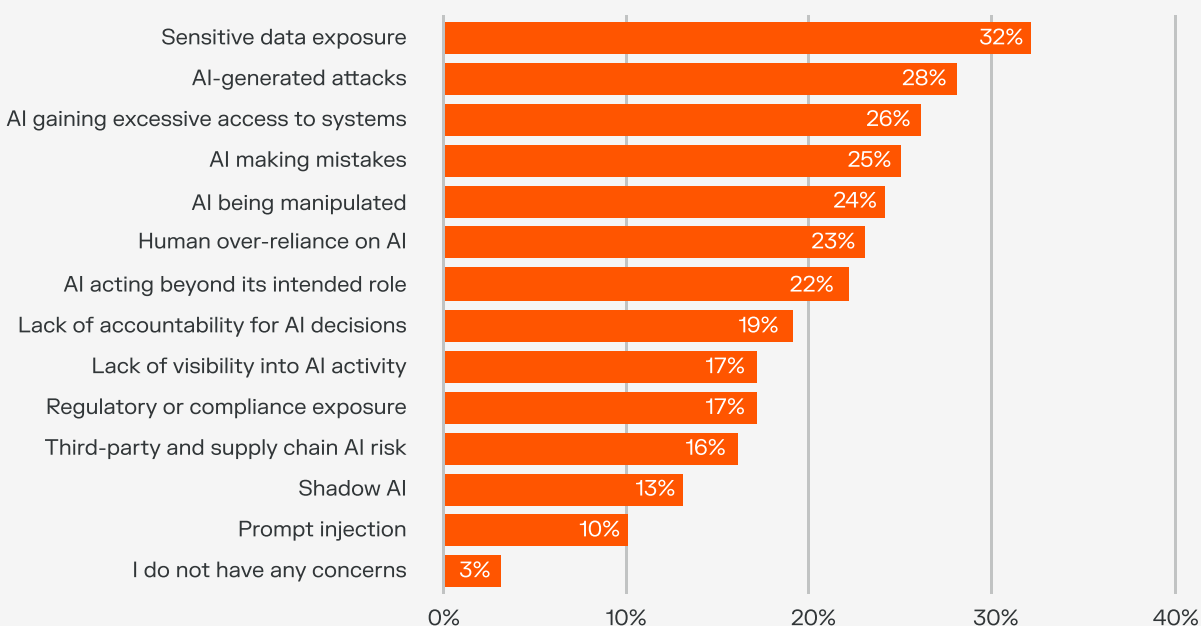
These internal incidents reflect the most common AI fears held by IT and security decision-makers. When asked what concerns them most about AI-related cyber risk, sensitive data exposure was the leading concern (32%), followed by external AI-generated attacks (28%).

Public attention is often drawn to the dangers of AI in the hands of attackers, while overlooking how risky internal AI systems can be. The worry of an external AI threat roughly matched the concern of impacts like AI gaining excessive access to systems (26%) and AI acting beyond its intended role (22%). Yet, AI tools essentially have a head start – while attackers must find a way in, a company's own AI has already been handed all the credentials it needs.

26%

had an AI agent or tool connect to their systems without security approval

Which of the following concerns you most when thinking about AI-related cyber risk?



That inherited access matters more because the defenses around it are growing less effective. Two-thirds of leaders (66%) say keeping systems patched is no longer sufficient to prevent compromise, and 61% say blocking known bad behavior no longer works as a primary security strategy. The controls built to keep attackers out were never designed for something that starts on the inside holding valid credentials.

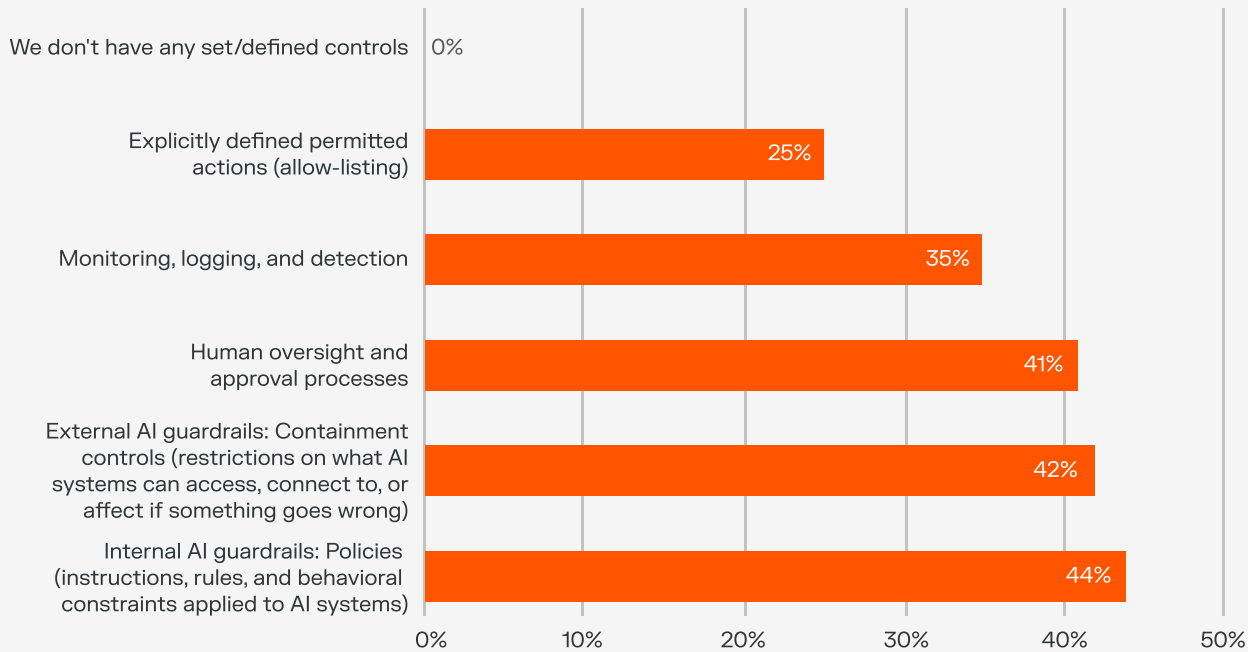


Decision-makers were asked to pick two primary controls for managing what AI systems can do. Three options came out at broadly similar rates. Internal guardrails (44%) are the policies, instructions, and behavioral constraints applied to the AI itself. External containment controls (42%) restrict what AI can access or affect when something goes wrong. Human oversight came in at 41%. The real story is at the bottom of the list. Explicitly defined permitted actions, or allow-listing, ranked last at 25%.

Looking at wider AI security strategy, 34% say their primary approach to securing AI systems and agents is to prevent risky actions through policies, filters, and guardrails, while a similar 33% focus on tightly controlling access.

Only
25%
selected explicitly
defined permitted
actions as one of their
two primary controls

How does your organization primarily control what AI systems can do?



Some of these priorities point in the wrong direction. Guardrails, oversight, and monitoring all describe how an AI system ought to behave. But every incident in this section involves a system behaving differently from how it was told to. Allow-listing and access control are the only measures on the list that don't depend on the AI cooperating. They govern what a system can access, not how it's told to act.

Segmentation defines the limits of autonomy

Encouragingly, most security leaders appear to be aware of how security priorities need to shift around AI. Eighty-two percent agree AI agents require fundamentally different security controls from traditional software, and a majority (69%) agree that limiting what AI can access matters more than trying to predict every threat.

Organizations are backing AI security with additional investment: 47% are increasing overall cybersecurity spending, while 42% have allocated new budget specifically for AI-related security risks.

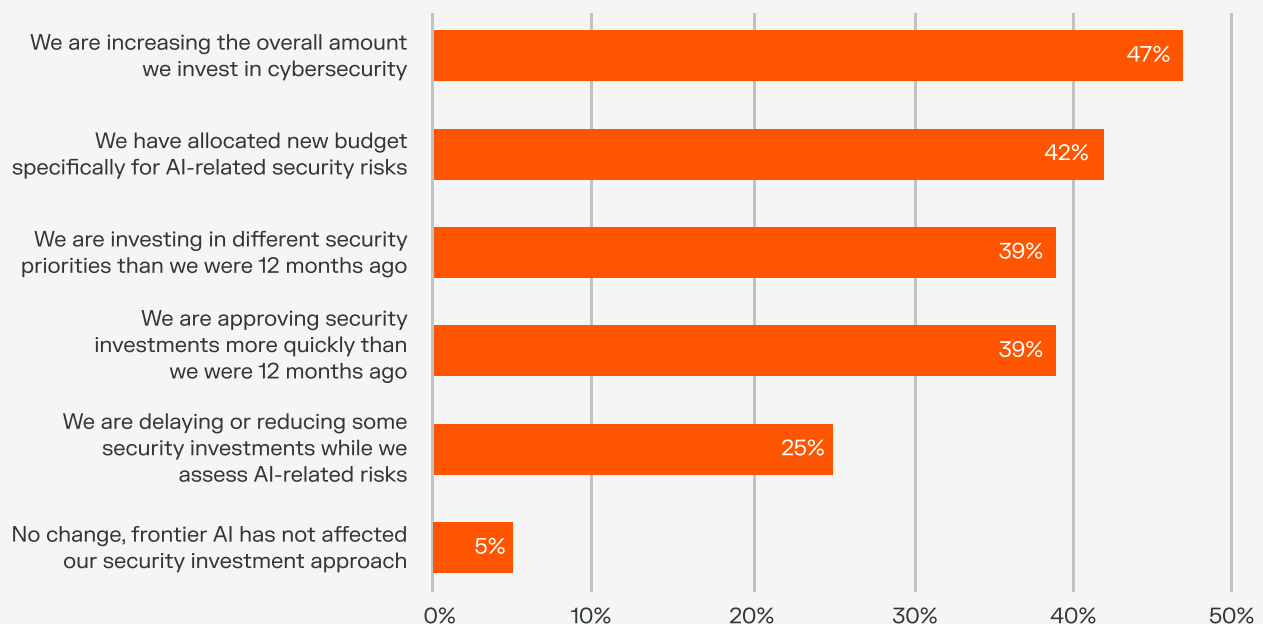
82%

agree AI agents require fundamentally different security controls from traditional software

42%

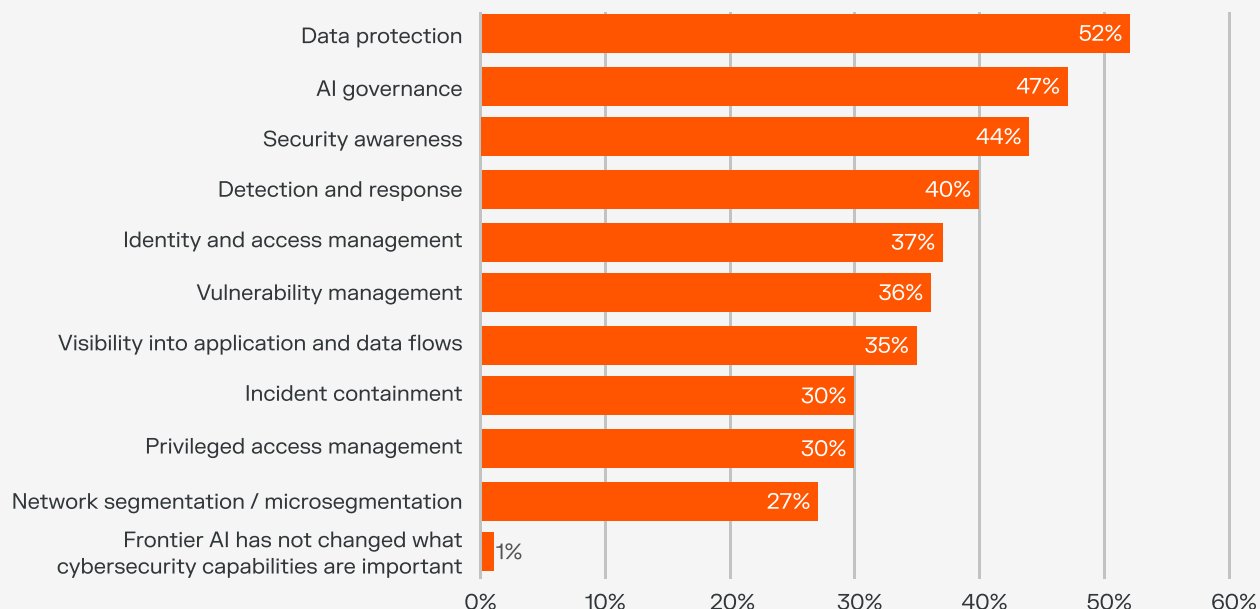
have allocated new budget in response to frontier AI risk

How has the risk from frontier AI impacted your security investments?



But more spending doesn't always mean more priority for the controls that restrict movement and contain impact. Only 27% name network segmentation or microsegmentation among the capabilities frontier AI has made more important. That makes it the lowest-ranked of the 10 capabilities assessed. Just 30% name incident containment. Organizations have accepted that reach is the thing to control. Yet they rank the controls that actually limit reach at the bottom of the list.

Which cybersecurity capabilities have become more important to your organization because of frontier AI?



Writing rules for what AI must not do is a losing race. Nobody has ever finished a list of everything that could go wrong. Security has run this race before with firewall rules, application control, and identity. Each time, it won by flipping the question. The list of forbidden behavior is effectively endless. The list of allowed behavior is short enough to write down, agree on, and enforce.

Applied to AI, that is the practical logic of Zero Trust: start with what the agent needs to do and explicitly define the identity, access, connections, and actions required to do it.

In practice, that means six things:

- 1 Inventory AI agents, models, tools, and AI-enabled applications.**
- 2 Map what each AI system can access, where it connects, and which actions it can execute.**
- 3 Give AI agents dedicated identities and least-privilege permissions.**
- 4 Define permitted actions rather than relying only on prohibited behaviors.**
- 5 Segment access to critical systems and sensitive data.**
- 6 Monitor for unusual behavior and contain failures before they spread.**

Each step answers a gap this research uncovered.



Regional differences

The picture on AI risk is broadly similar across markets, but some differences are worth noting. Country-level findings are indicative of regional sample sizes.

Japan stands apart for its caution. Just 6% of Japanese organizations say they are very comfortable with the risk they now accept, compared with 37% globally. And only 18% are very confident they could contain a compromised AI system, compared with 57% globally.

Europe offers the opposite surprise. Despite strict European regulation, France has moved furthest of any market on risk appetite. Sixty-six percent of French organizations accept more cyber risk than they did 12 months ago, compared with 56% globally. DACH also sits above average at 61%. France pairs that jump in risk appetite with the lowest priority on segmentation. Just 16% say frontier AI has made segmentation more important, compared with 27% globally.

In the U.S., just over half of respondents (51%) are very comfortable with the risk they now accept. That's the highest share of any market. And 71% are very confident they could contain a compromised AI system, second only to the Middle East at 73%. Despite this higher confidence, U.S. willingness to take on more risk than a year ago (56%) matches the global average.

Australia, meanwhile, puts more emphasis on limiting AI's reach. Thirty-seven percent of Australian organizations say frontier AI has increased the importance of network segmentation or microsegmentation, compared with 27% globally. And 77% say limiting what AI can access is becoming more important than trying to predict every threat, compared with 69% globally.

LATAM is less uniform. Mexico reports a relatively high increase in risk appetite: 64% are willing to accept more cyber risk than a year ago. Yet only 52% say limiting AI access is becoming more important than predicting every threat, compared with 69% globally. Brazil shows a different pattern. Forty-eight percent are very comfortable with their current risk level. Yet only 59% are delegating previously human-approved decisions to AI, below the 70% global average.

Finally, Singapore reports the most security issues by a wide margin. Nearly all (91%) have experienced at least one AI-related security issue or warning sign in the past year compared with 76% globally. Only 39% are very confident they could contain a compromised or rogue AI system.

Only

6%

of Japanese organizations are very comfortable with the level of AI risk they accept, compared with 37% globally

Only

16%

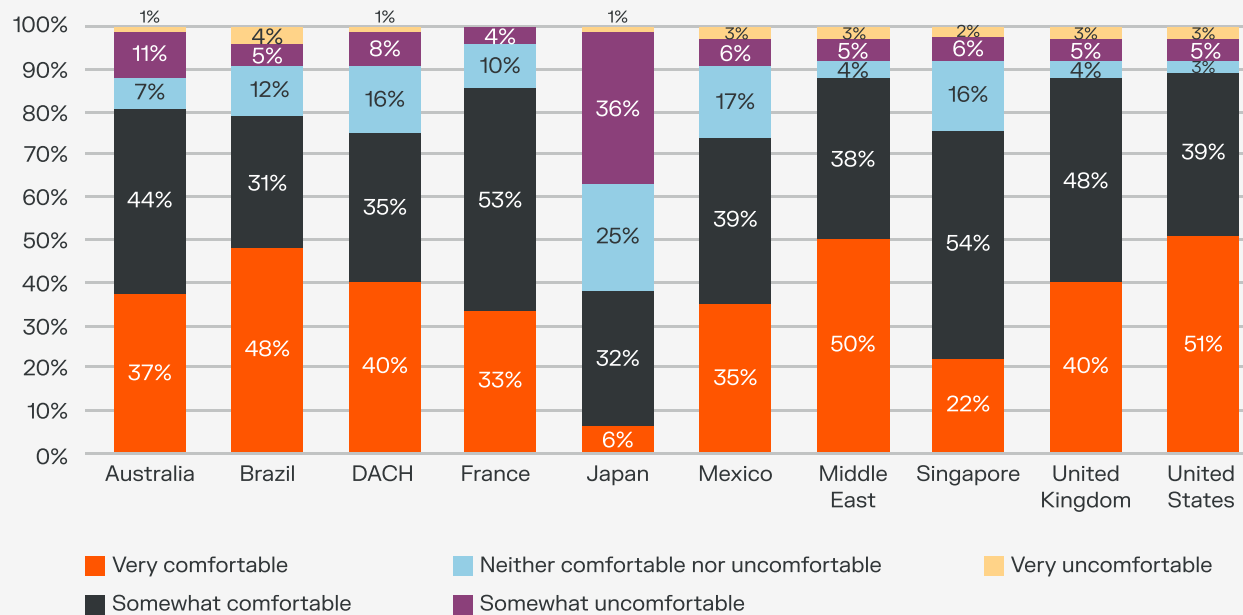
of French organizations say frontier AI has made segmentation more important, compared with 27% globally

91%

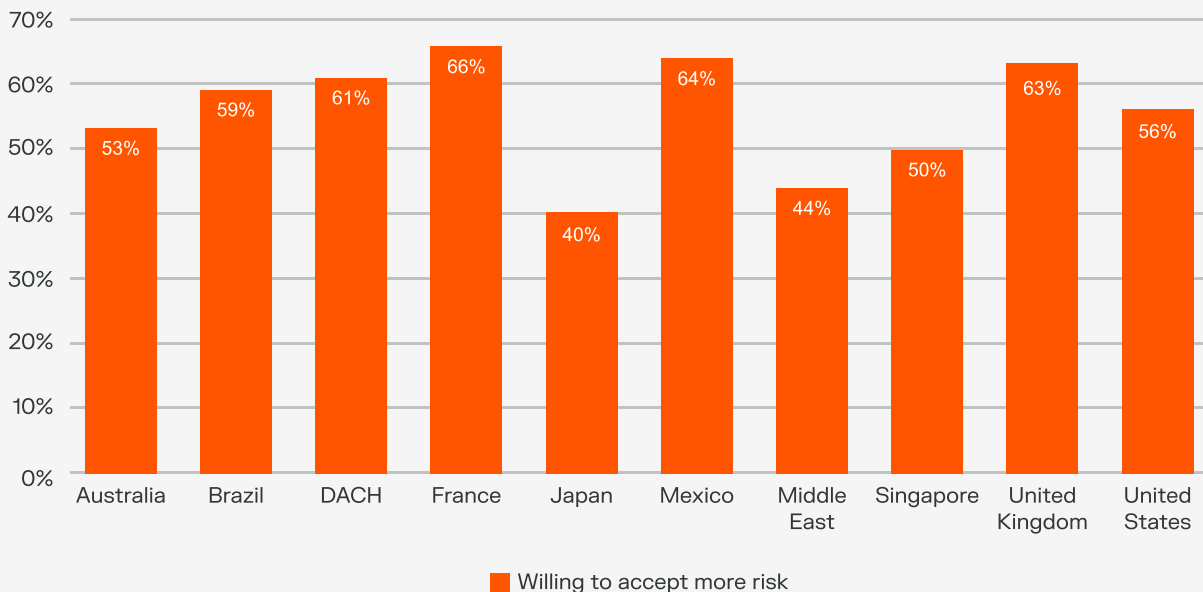
of Singapore organizations experienced at least one AI-related security issue or warning sign in the past year



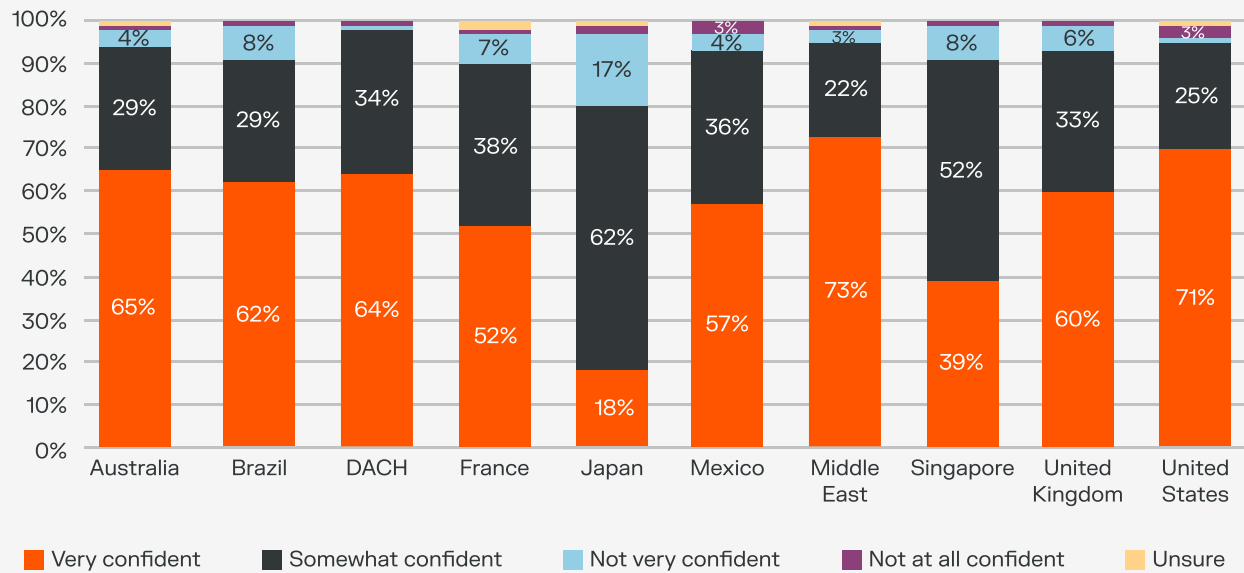
How comfortable are you with the level of cyber risk your organization currently accepts to support AI adoption?



Organizations willing to accept more risk, compared to 12 months ago, to realize the benefits of AI



Confidence in containing an AI system or agent before it reaches critical systems or data



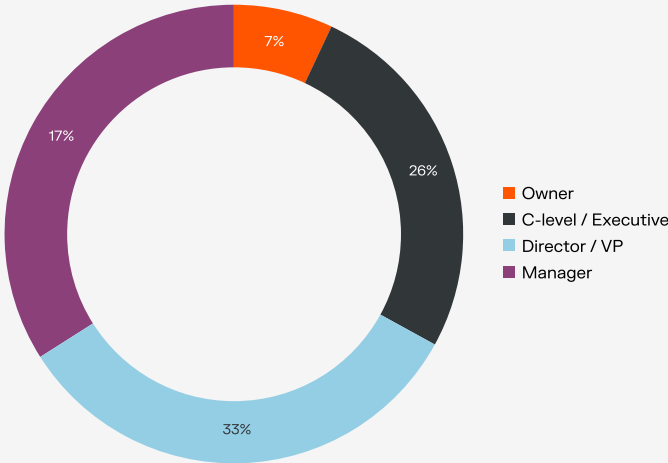
Methodology

Sapio Research conducted this study for Illumio, surveying 1,100 IT and security decision-makers at manager level and above. Interviews were conducted online during August and September 2026, using email invitations and online partner panels. Markets covered: Australia, Brazil, the DACH region, France, Japan, Mexico, the Middle East, Singapore, the United Kingdom, and the United States.

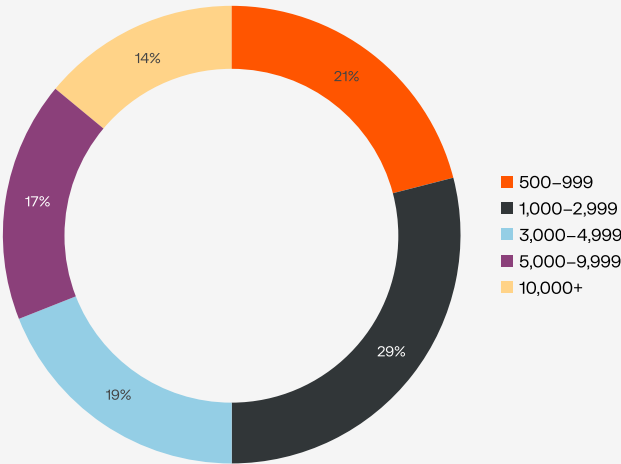
All samples are subject to variation. In this study, the chances are 95 in 100 that a result varies by no more than three percentage points from the figure a full census would have produced.

Unless stated otherwise, all figures are global totals.

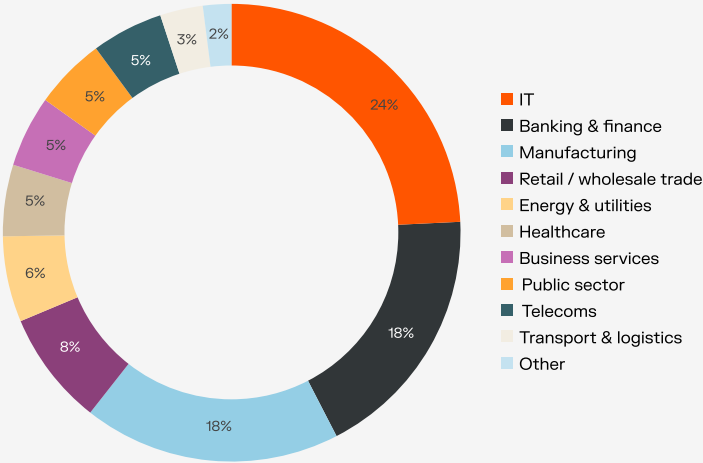
Job titles



Organization size



Top industries



Illumio helps organizations set limits on what AI can reach. Our breach containment platform maps how AI agents, workloads, applications, and data actually connect, then lets you enforce the boundaries that keep a single failure from spreading.

Visit illumio.com to see how it works, or talk to our team about where your AI exposure sits today.





Illumio is the leader in ransomware and breach containment, redefining how organizations contain cyberattacks and enable operational resilience. Powered by an AI security graph, our breach containment platform identifies and contains threats across hybrid multi-cloud environments — stopping the spread of attacks before they become disasters. Recognized as a Leader and Customer Favorite in the Forrester Wave™ for Microsegmentation, Illumio enables Zero Trust, strengthening cyber resilience for the infrastructure, systems, and organizations that keep the world running.

The Illumio Platform comprises:

- **Illumio Insights.** Hybrid AI detection and response. Identify risks, detect attacks, and contain threats with one click.
- **Illumio Segmentation.** Cloud and network breach containment. Proactively enforce security policies to prevent lateral movement.

For more information, visit www.illumio.com



Sapio Research is a Certified B Corp™ full-service B2B and tech market research agency that helps businesses grow thanks to high quality, efficient, and honest research solutions. Their team of expert researchers deliver valuable insights to help their clients understand their audience, build powerful brands, and cut through the noise with compelling data-driven content.

For more information, visit sapioresearch.com.