knowbe4

May 2026

From Agentic Risk to Human Wins

Building a Culture of Security in the Era of Agentic AI

The Rules Just Changed.

For decades, cybersecurity was often viewed as a technology problem with a human variable. Train the people. Patch the systems. Repeat. But agentic AI has rewritten that equation entirely. The question is no longer how we protect humans from threats. It's how humans and AI agents work together, build trust and make high-stakes decisions in real time. In that world, agentic AI is simultaneously a teammate, a decision-maker and an attack surface waiting to be exploited.

Most organizations aren't ready for that. But this research shows that the ones moving in the right direction share something important: they've stopped treating security as a function and started treating it as a culture. Secure behaviors are woven into daily work. Employees don't just comply. They contribute. These organizations have stopped chasing the threat and started shaping the conditions around it.

The gap between those organizations and the rest is closing fast. The cost of being on the wrong side is rising just as quickly.

At KnowBe4, we've always held that humans are not the weakest link. They're the most powerful defense you have. Full stop. That mindset and mission now extend to the AI agents working alongside them. Our mission is to help organizations build the culture and the trust that makes humans and agentic AI stronger together. Every person in your workforce can be an active, adaptive participant in that future. We're here to make sure they are.

The organizations that will define the next era of cybersecurity aren't waiting for a better threat model. They're building better humans, better cultures and better partnerships with the AI working beside them. This report shows you where to start.

This research doesn't predict the future. It describes the organizations already living in it. The question isn't whether you'll get there. It's whether you get there first.

Let's build a future where we move from agentic AI risk to human wins... together.

Perry Carpenter

Perry Carpenter
Chief Deception Strategist

What's Inside

Introduction

Cybersecurity has always been more than a technology challenge. It sits at the point where people and technology meet. This is where human behavior often drives risk in ways tools alone cannot fix. AI accelerates these challenges and reshapes how risk shows up in everyday work.

The role of human behavior in cyber risk has always been recognized by cybersecurity leaders. Yet breaches persist because it is difficult to apply this understanding with real-world conditions. Risk continues to emerge through everyday decisions made under pressure, distraction and uncertainty.

The challenge is now intensifying as the workforce itself changes. Employees no longer work alone. They work alongside agentic AI tools and autonomous agents that generate content, take actions and interact with sensitive data. This has created a new, more complex environment where human and agentic AI behaviors (and risk postures) intersect.

This research explores the shift. We surveyed 800 cybersecurity leaders and 3,200 employees across the globe. This is what they told us about human and agent risk, and whether they think they are ready for what comes next:

58%
of surveyed cybersecurity leaders say that **everyday mistakes are a key driver** of cybersecurity risks

42%
of cybersecurity leaders identify **AI-enabled attacks as a top driver** of future human-related cyber security risks

36%
of cybersecurity leaders report **significant improvement in managing human risk** within the last 12 months

97%
of cybersecurity leaders are **using metrics to track human-related cybersecurity risks**, yet only 61% say these fully support their efforts

42%
of cybersecurity leaders strongly believe that **security awareness alone drives lasting behavior change**

89%
of employees agree they **feel safe when reporting security mistakes**

19% of cybersecurity leaders report their organization **operates an integrated and culture-embedded approach** to managing human-related cybersecurity risk

- Those who have adopted an integrated and culture-embedded approach are more likely to say security awareness activities build trust between the IT team and employees, provide clear guidance for reporting and take a coaching-led approach to phishing simulations.

Section 1

The Reality of Risk in the Integrated Digital Workforce

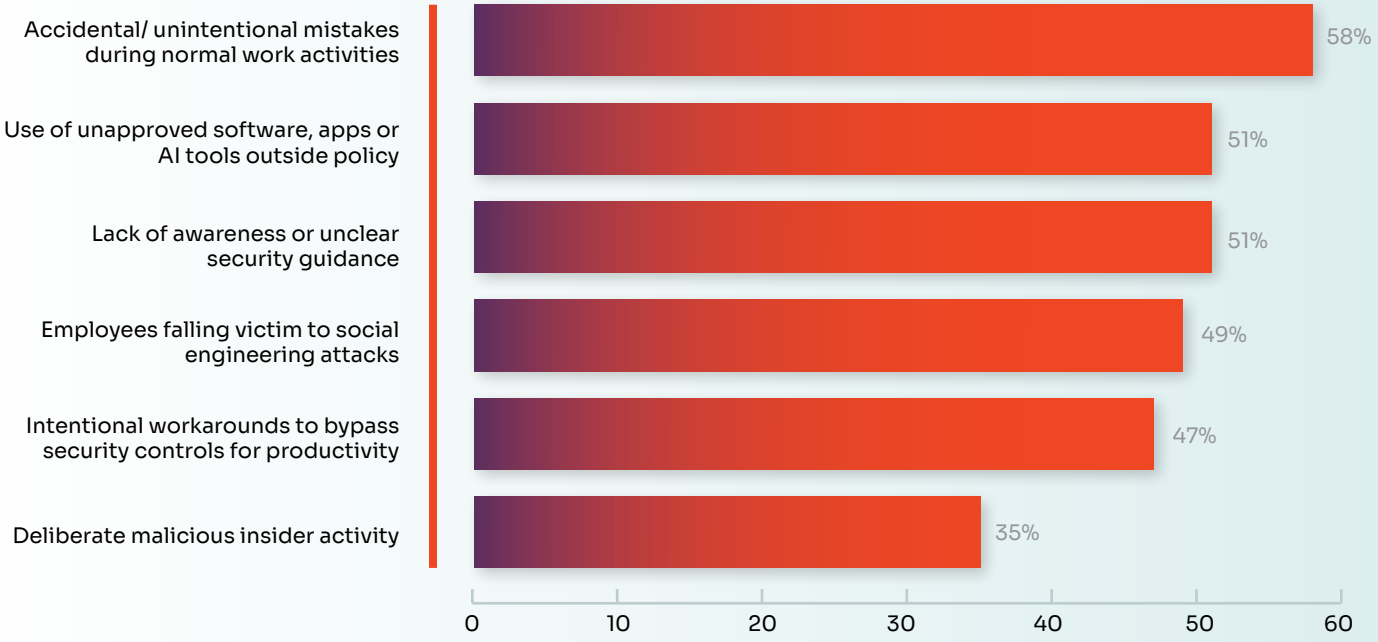
Organizations are evolving their approaches to workforce security and human-related cyber risk, but the nature of risk itself is also changing. The way we work has shifted, with employees now having to make faster decisions, navigate complex systems and work alongside agentic and non-agentic AI tools that influence decisions and actions. In this environment, human and agent mistakes are both common and consequential, increasing risk in everyday work behaviors.

> "Be proactive and cautious, do not over rely on AI, always have [a] human overseeing things, double checking things. We have got robust processes and procedures for checking."
>
> Senior Manager, Healthcare and Pharmaceuticals, 5000+ Organizations

Risk is emerging through everyday work, not isolated incidents

Almost six in 10 cybersecurity leaders (58%) report that mistakes during everyday work have had the greatest impact on their organization's cybersecurity in the past 12 months.

Types of human-related behaviors with greatest impact on cybersecurity in the last 12 months



Q7. In the past 12 months, which of the following types of human-related behavior has had the greatest impact on your organization's cybersecurity? (Base: Cybersecurity Decision Makers, 800)

Mistakes, by their very nature, are not malicious; they are the reality of day-to-day work. As workflows become more complex and productivity pressures increase, employees are increasingly required to balance speed with security. In fact, cybersecurity leaders identify pressure to prioritize speed over security as one of the leading drivers of rising human risk in the next 12 months, just behind the increasing complexity of tools and AI-driven attacks.

55% of employees agree that they may know the safe action to take, but time pressure or distractions can lead to mistakes

Q4. Thinking about your day-to-day work, to what extent do you agree or disagree with the following statements? 'Even when I know the safe action to take, time pressure or distractions can lead me to make a 'security mistake' (Base: Employees, 3200)

This reflects a broader challenge. As organizations provide employees with LLMs and agentic AI tools to move faster, they may not always be putting relevant guardrails in place. Without these controls, an AI-enabled workforce will lead to greater risks instead.

Over half (55%) of employees themselves also admit that even when they know the safe action to take, time pressure or distractions could lead them to making a security mistake. This highlights two key realities: how easily intentions can break down under pressure even when employees understand the risks; and the gap between awareness and behavior. Awareness is a state of mind, but behavior is what ultimately determines security outcomes. Closing that gap requires more than communication alone. It requires organizations to actively shape employee behavior by embedding security into everyday workflows and build a culture where safe actions are the default. Only an effective digital workforce security program provides the bridge needed to close this gap.

This gap between awareness and behavior becomes further pronounced as agentic AI is integrated into everyday workflows, often without users consciously recognizing it. Employees are turning to AI to draft emails, summarize meetings, analyze spreadsheets, organize files and generate reports, while developers use AI-assisted coding tools to accelerate software delivery. The use of AI has shifted from being an optional tool to an invisible automated agentic layer that continuously augments how work gets done... and it does so at machine speed and scale.

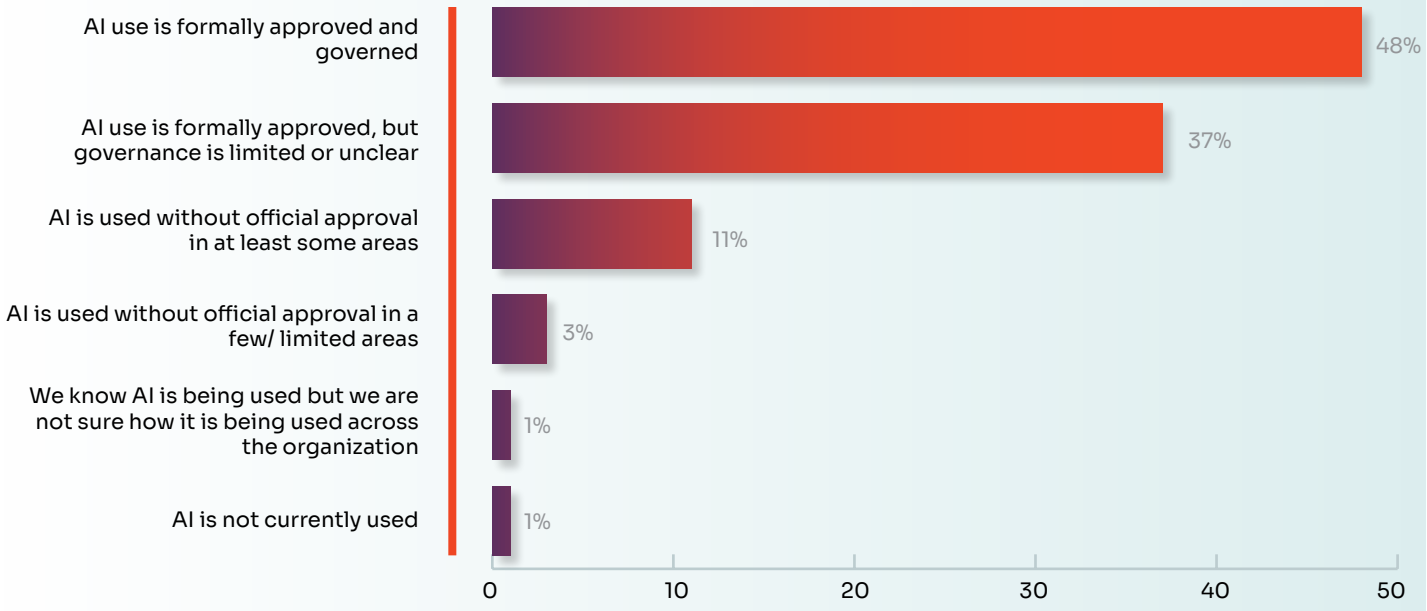
Agentic AI is increasing exposure faster than its being controlled

Agentic AI tools are now widely embedded in day-to-day work, with 58% of cybersecurity leaders reporting that agents are already taking actions within organizational workflows. At the same time, a lack of oversight and governance is leaving organizations exposed. **Only 48% describe their AI use as formally approved** and governed, while a further 37% say while AI use is formally approved, governance is limited or unclear.

58% have AI tools/ AI agents taking actions autonomously in multiple workflows, including 17% with limited human oversight

Q18. Which of the following best describes how AI tools or AI agents are used in your organization's workflows today? (Base: Cybersecurity Decision Makers whose organization is using AI, 795)

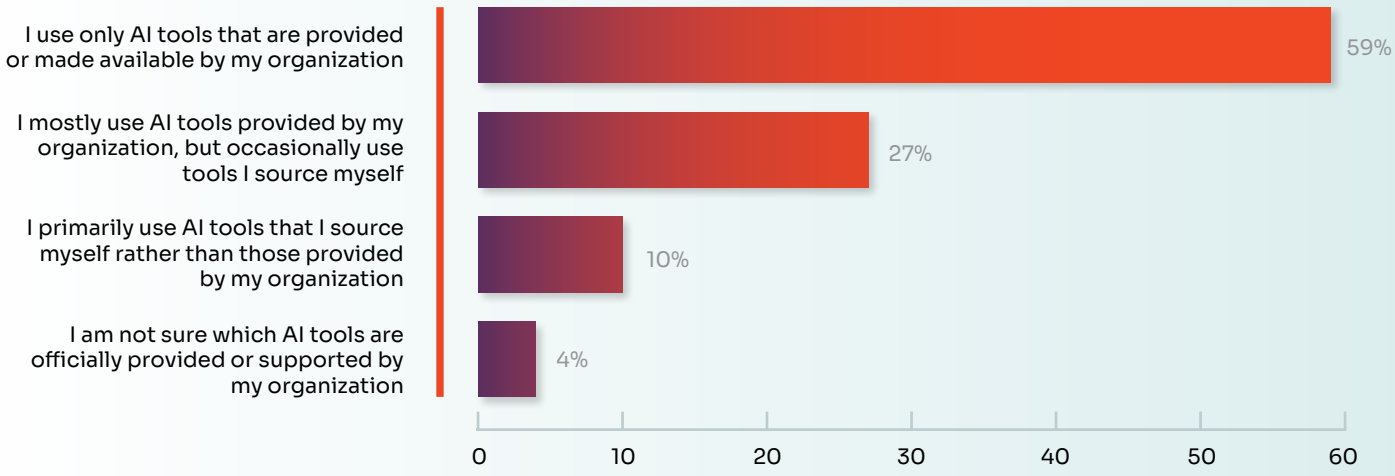
How AI is being used and governed by organizations across their businesses



Q17. Which of the following best describes your organization's use of AI tools (including generative AI) across the business?
(Base: Cybersecurity Decision Makers, 800)

Over a third of employees also reported that they commonly source their own agentic AI tools where official options are unavailable or restrictive.

How employees use AI in their day-to-day work



Q6. Which of the following best describes how you use AI in your day-to-day work?
(Base: Employees who say they are using AI tools at work, 2637)

Over half (51%) of cybersecurity leaders also believe that the use of "Shadow AI" or IT tools has had a great impact on their organization's cybersecurity over the past 12 months. This impact reflects a shift where these unsanctioned tools are no longer just unmanaged software, but act as "shadow employees," autonomously performing tasks without oversight.

Organizations struggle to enforce consistent policies and oversight, with nearly half (47%) reporting the safe use of AI (both tools and agents) as a key challenge for their organization. As a result, risk is shifting toward how chatbots and agentic AI are used in practice, including what employees share, how they interpret outputs and whether they verify recommendations. Encouraging good practice and behavior around agentic AI use is just as important as securing the technologies themselves.

51% believe the use of unapproved software, apps or AI tools has had a great impact on their organization's cybersecurity in the past 12 months

Q7. In the past 12 months, which of the following types of human-related behavior has had the greatest impact on your organization's cybersecurity?
(Base: Cybersecurity Decision Makers, 800)

AI Tool Vs AI Agent

- AI tools are typically reactive and perform specific tasks when prompted by a user, such as generating text, summarizing documents or analyzing data. An AI tool will generally operate within a defined scope under direct human control. AI agents are more autonomous and proactive: they can plan, make decisions and execute multi-step actions on a user's behalf, often interacting with connected systems like email, calendars, files and business applications. Because agents can monitor conditions, trigger workflows and adapt their behavior over time, they enable greater productivity but also introduce more complexity and risk if not properly managed.

Human-AI interaction is a key point of vulnerability

The threat landscape is evolving as employees and agentic AI work together. AI-enabled attacks, including highly personalized phishing and deepfake content, are becoming more convincing and harder to detect. More than four in five (86%) employees say deepfake content is now so realistic that it is harder to know what to trust. At the same time, 64% believe they could still be tricked by these attacks. This leaves a notable minority who do not think they are at risk, suggesting a potential false sense of security as these threats become more advanced.

Phishing remains a key driver of risk. Nearly six in 10 (59%) employees identify phishing or impersonation emails as a main cause of human-related cyber risk, with 26% perceiving it as the number one threat.

At the organizational level, 42% of cybersecurity leaders identify AI-driven attacks as a key factor driving increased human-related cyber risk in the next 12 months. Despite this, preparedness is limited and while many organizations feel generally ready to manage emerging risks, fewer than half (48%) say they are *very well prepared* to handle unexpected or emerging AI-related threats.

This disconnect between risk and preparation presents an opportunity for organizations to advance their approach to proactive risk management strategy. As agentic AI adoption accelerates and threats become more sophisticated, the focus is moving toward strengthening governance, shaping behavior and improving readiness.

Risk has always existed at the intersection between IT systems and its users, but agentic AI is now accelerating this complex dynamic. This makes the human–AI interaction layer a critical area to secure.

86% of employees say deepfake content is now so realistic it is harder to know what to trust

64% of employees say it is possible they could be tricked by AI-enabled attacks

Q10. To what extent do you agree or disagree with the following statements about AI-enabled cyber threats? 'Deepfake voice or video has become so realistic that it is now harder to know what to trust' / 'It is possible that I could be tricked by an AI-enabled scam or threat at work, such as a deepfake call or message' (Base: Employees, 3200)

42% believe they are very well prepared to manage emerging and unexpected human-related risks in the next 12 months

48% believe they are very well prepared to manage emerging and unexpected AI-related risks in the next 12 months

Q24. To what extent would you say your organization is prepared to manage unexpected or emerging human- and AI-related cybersecurity risks over the next 12 months? (Base: Cybersecurity Decision Makers, 800)

Section 2

The Maturity of Human and Agent Risk Management

As day-to-day work becomes increasingly shaped by agentic AI, organizations are adapting their approach to workforce security and human-related cyber risk. It is now becoming a central pillar of cybersecurity strategy. However, maturity remains uneven, with confidence often outpacing actual capability.

Most organizations report progress, with nine in 10 (90%) saying their ability to manage human-related cyber risk has improved over the past 12 months. However, only 36% describe this improvement as significant, suggesting that while many organizations feel they are making progress, fewer are seeing real changes in behavior or culture.

36% believe their ability to manage human cyber risk has significantly improved in the last 12 months

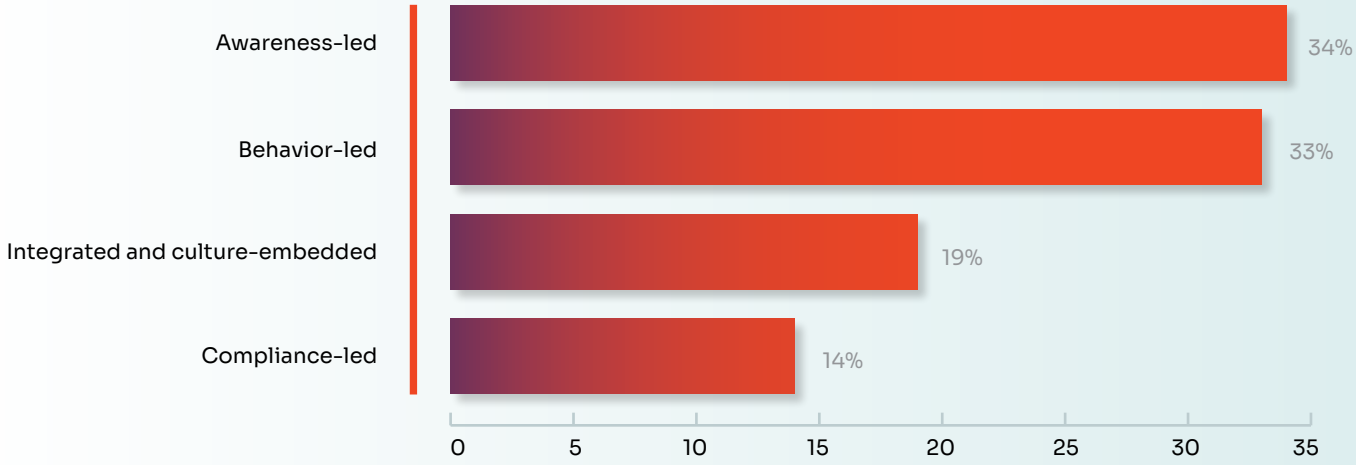
Q3. How has your organization's ability to manage human-related cybersecurity risk changed in the last 12 months? (Base: Cybersecurity Decision Makers, 800)

While 97% say their assessment of human risk is supported by metrics or evidence, fewer demonstrate the level of visibility required to translate those metrics into actionable insight. Measurement exists, but it is often fragmented, inconsistent or disconnected from real-world behavior.

This gap becomes more apparent when looking at how organizations are managing risk in practice. Many organizations remain anchored in more traditional security awareness approaches, with just over a third (34%) relying on awareness-led strategies that focus on providing knowledge and hoping it will translate into behavioral change.

However, earlier findings showed that 55% of employees admit to being at risk of making a security mistake due to time pressures or distractions, despite knowing the safe course of action to take. This reinforces that awareness alone is not enough. Awareness does not always translate into action, particularly when employees are under pressure or relying on default behaviors. Additionally, only 42% of cybersecurity leaders strongly believe that awareness training alone can drive lasting employee behavior change.

How organizations are currently reducing human-related cybersecurity risk



Q5. Which one of the following best describes how your organization currently reduces human-related cybersecurity risk? (Base: Cybersecurity Decision Makers, 800)

This reliance on awareness-led approaches highlights a fundamental tension at the heart of workforce security and human-related cyber risk. Organizations understand that behavior matters, but many are still operating as if awareness is sufficient. This mindset is reflected in the fact that around four in 10 organizations cite driving employee engagement, ownership or buy-in (42%) and encouraging secure behavior (40%) as a key challenge in managing cyber risk.

In reality, behavior is shaped by a combination of context, employee's motivation and their ability to act. Employees may know the correct action to take, but under pressure, speed or cognitive overload, they often act differently. This means that reducing risk is not only about improving awareness but also changing the conditions in which decisions are made.

The "gold standard" is a fully integrated and culture-embedded approach to workforce security, where organizations combine awareness, behavior and compliance, supported by strong leadership, clear ownership and effective measurement. Yet only 19% of organizations report this level of integration.

Organizations that adopt this more mature approach are fundamentally different to others. Rather than relying on awareness alone, they are doing the following:

- Embed security into everyday workflows and make safe actions the default choice
- Reinforce behavior through systems and real-time guidance, not just training
- Combine awareness, behavior and governance into a coordinated approach

Around half (49%) of organizations believe they are building secure behavior into workflows. This highlights a clear perception gap. For most, progress is underway, but not yet at the level required to deliver consistent, embedded outcomes.

Workforce security has always been more than awareness and compliance. Yet in practice, many organizations may have a limited understanding and focus on training and policy rather than fully addressing other factors such as behavior, culture and integration. As a result, many organizations have not yet reached a well-defined approach to addressing human-related cyber risk, and execution does not always align with intent.

Section 3

Trust, Culture and the Drivers of Behavior

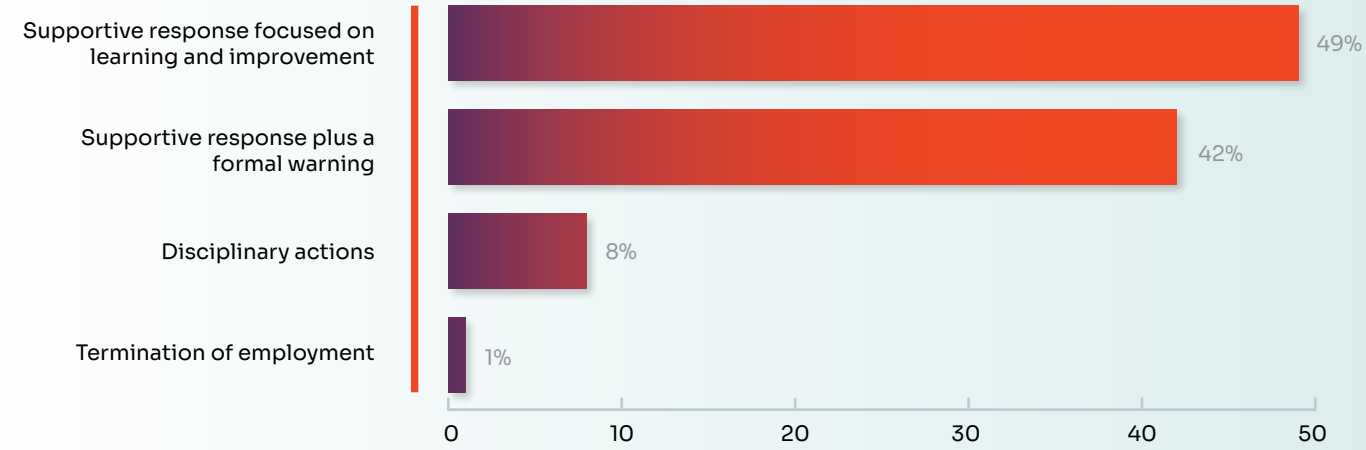
Organizations are increasingly recognizing the importance of trust and psychological safety in driving secure outcomes. Most respondents report that employees feel safe reporting mistakes to their IT team, with 93% of cybersecurity leaders and 89% of employees aligning with this perception.

Cybersecurity often relied on punishment or compliance-driven approaches, where mistakes were met with discipline. While this may enforce rules, it may also discourage transparency.
As a result, organizations need to build a culture of confidence, where employees feel supported to report mistakes or issues and trust that their actions will be treated with respect and consideration.

From enforcement to coaching-led security approaches

Organizations are moving toward more supportive, coaching-led approaches. Nearly half (49%) now take this response when employees make security "mistakes," while 63% have implemented learning-led phishing simulations that are designed to educate and coach rather than catch employees out (although for 24%, there are formal consequences issued).

How organizations typically respond when an employee makes an inadvertent cybersecurity mistake

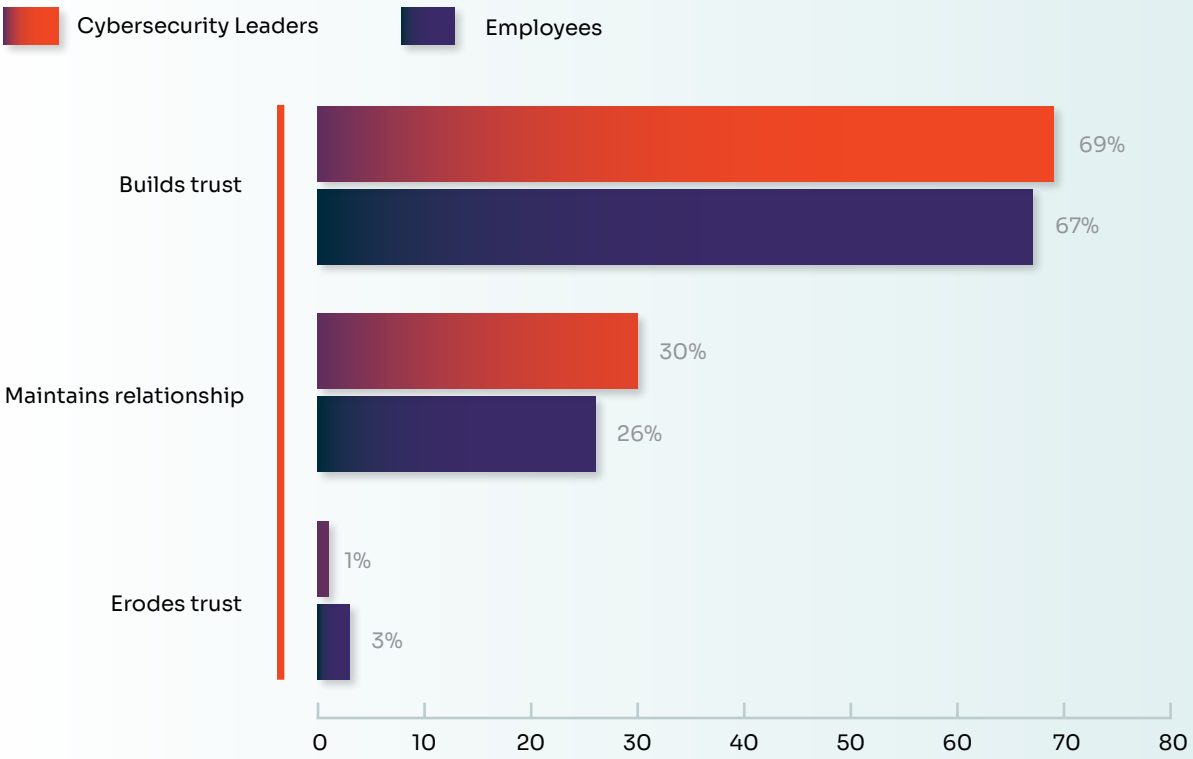


Q12. How does your organization typically respond when an employee makes an inadvertent cybersecurity mistake? (Base: Cybersecurity Decision Makers, 800)

Organizations have built safe environments for employees, but this varies across regions

When employees trust security teams and feel supported, they are more likely to act as active participants in defense. This is reflected in the broader relationship between employees and security functions, with 91% viewing IT and security teams as supportive partners. Additionally, around two thirds of employees and cybersecurity leaders (69% and 67% respectively) believe that security awareness activities help build trust between IT teams and employees.

Influence of cybersecurity awareness activities on IT-employee relationships



Q6. In your view, how do your organization's cybersecurity awareness activities influence the relationship between the IT/ security team and your employees? (Base: Cybersecurity Decision Makers, 800) / Q3. In your view, how do your organization's security awareness activities influence the relationship between employees and your IT/ security team? (Base: Employees, 3200)

However, this progress is not consistent across regions. In the APJ (Asia, Pacific, Japan) region, for example, confidence in reporting and trust in security culture is notably lower compared to other regions.

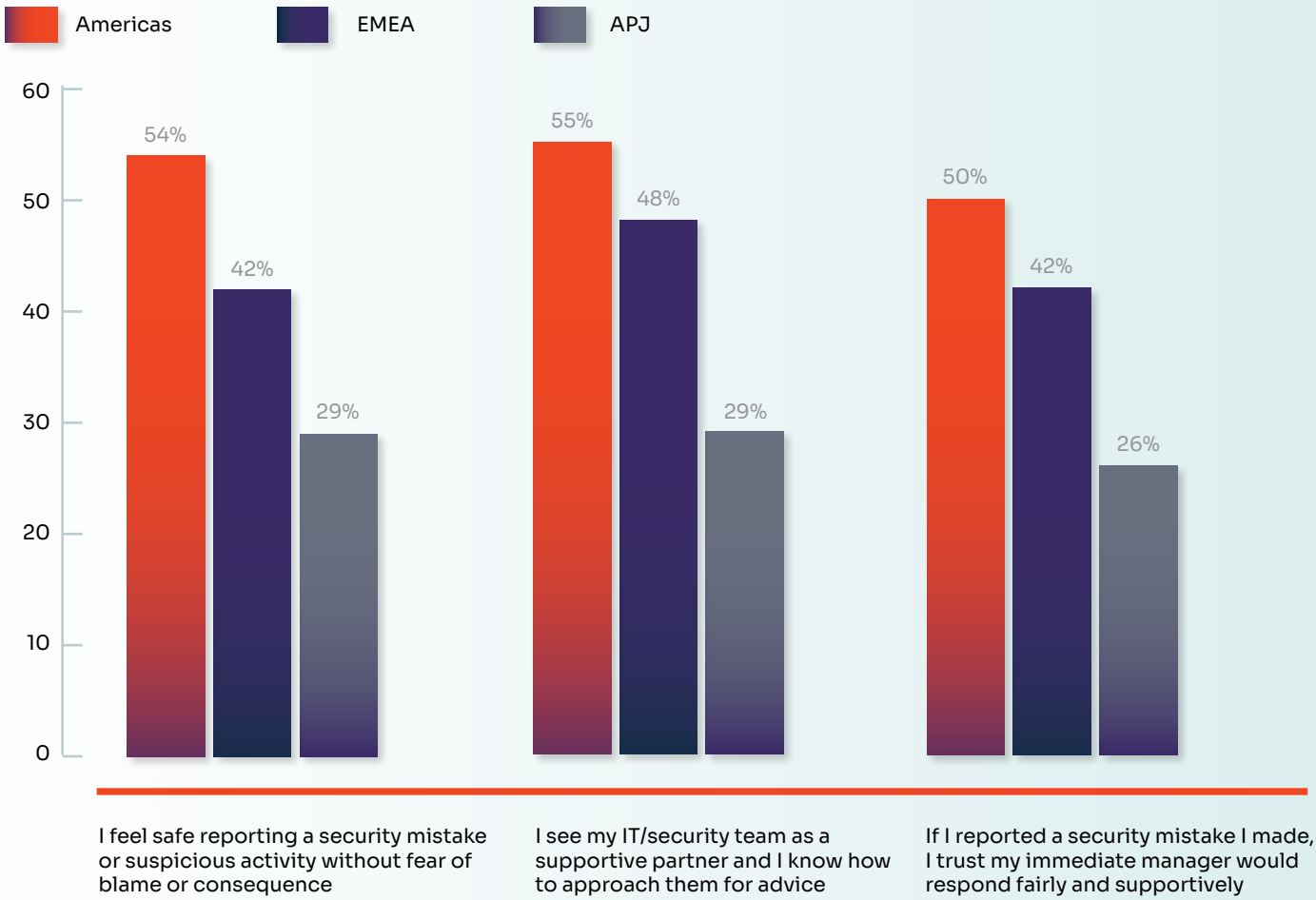
Nearly three in 10 (29%) employees in APJ feel safe reporting a security mistake, compared to 54% in the Americas and 42% in EMEA. This varies further within the region with perceived safety being highest in Australia and New Zealand (43%) and lower in Japan (21%) and Singapore (27%).

A similar pattern is seen in how APJ employees perceive their IT teams. Only 29% strongly believe that their IT team is a supportive partner, compared to 55% in the Americas and 48% in EMEA. There are also clear differences within the region with lower levels of trust in Japan (24%) and Singapore (28%). Trust in immediate managers follows a similar pattern. Only a quarter (26%) of employees in APJ believe their manager would respond fairly and supportively when they report a security mistake This drops even further in Japan with only two in ten (20%) saying this.

These differences show how much culture can shape security behavior. Risks are less likely to be reported early when employees do not feel safe speaking up. Even the best tools can be limited in these situations as they rely on people to report issues immediately.

The research therefore reinforces that security is not just about policies or tools. It depends on trust, culture and on how people act in their jobs on a daily basis.

Regional differences for confidence in reporting and trust in security culture (% strongly agree)



Q4. Thinking about your day-to-day work, to what extent do you agree or disagree with the following statements? (Statements: I feel safe reporting a security mistake or suspicious activity without fear of blame or consequence, I see my IT/ security team as a supportive partner and I know how to approach them for advice, If I reported a security mistake I made, I trust my immediate manager would respond fairly and supportively)
(Base: Employees who strongly agree with statements, 3200)

Not all approaches deliver the same outcomes

Organizations that foster trust, reinforce positive behavior and create safe reporting environments are better positioned to translate awareness into action. This is reflected in more mature, **integrated and culture-embedded** cybersecurity approaches.

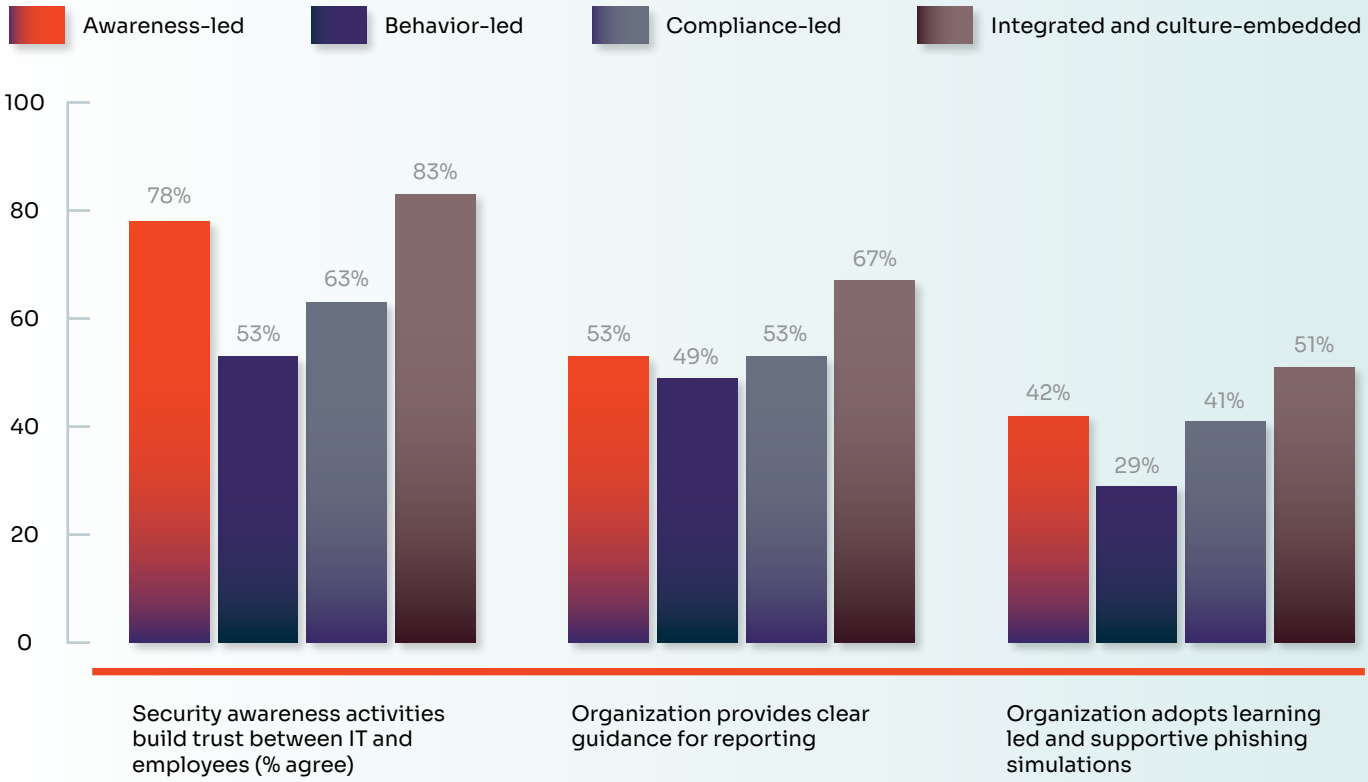
Organizations taking an integrated and culture-embedded approach combine awareness, behavior and compliance, supported by strong leadership, clear ownership and effective measurement. These organizations outperform others in certain key areas. For example, over eight in 10 (83%) say security awareness activities build trust between the IT team and employees, 67% provide clear guidance for reporting and 51% take a coaching-led approach to phishing simulations.

For those taking **awareness, compliance or behavior-led** approaches, the outcomes are less consistent and weaker than those who are using an **integrated and culture-embedded** approach, highlighting how important this approach is to the effectiveness of the security programs and culture in place within organizations.

What does this mean for how organizations approach human risk?

This difference in outcomes marks a broader shift in how organizations approach human risk. Rather than focusing only on preventing mistakes, more mature organizations are creating environments where employees are encouraged to act, report and improve. The result is employees who can support and enhance the overall security of the business.

How different approaches have an impact on various security culture metrics



Q6. In your view, how do your organization's cyber security awareness activities influence the relationship between the IT/ security team and your employees?/ Q14. Which of the following, if any, does your organization have in place to encourage employees to report suspicious cybersecurity activity or human-related cybersecurity mistakes? / Q13. Which one of the following best describes your organization's approach to phishing simulations or security testing of employees? AND cut by Q5. Which one of the following best describes how your organization currently reduces human-related cybersecurity risk? (Base; Cybersecurity Decision Makers, 268 Awareness-led, 264 Behavior-led, 115 Compliance-led, 152 Integrated and culture-embedded)

Conclusion: From Managing Risk to Enabling Resilience

The goal was never zero mistakes. It was always resilience. Organizations are building toward that now. Forward leaning organizations are working to combine the right tools, agentic AI and a culture where employees feel safe to speak up when something goes wrong. Technology matters. Culture matters more. Together, they create the conditions where human and agentic AI risk becomes human strength.

Technology, particularly agentic AI, is evolving faster than people can adapt. As such, three critical gaps define the current landscape:

1 Confidence and capability

Organizations believe they are improving, but measurement and visibility remain limited.

2 Awareness and behavior

Organizations are building awareness, but this does not guarantee employees will act accordingly in practice.

3 Agentic AI adoption and control

AI is already embedded in workflows, but governance and oversight are still catching up.

This is where the concept of “Human Wins” becomes paramount. Organizations that excel in workforce security will be those that:

- Design systems that guide behavior rather than just inform it
- Build supportive cultures that encourage reporting and learning
- Shift from tracking failures to reinforcing positive actions
- Extend a security-first mindset across both humans and AI agents

In this new reality, the digital workforce (comprising both employees and AI agents) becomes a critical layer in defending organizations from cyber risks. As work continues to change, organizations that bring together awareness, behavior and culture, and keep a clear view of their risk, can reduce exposure and build lasting resilience.

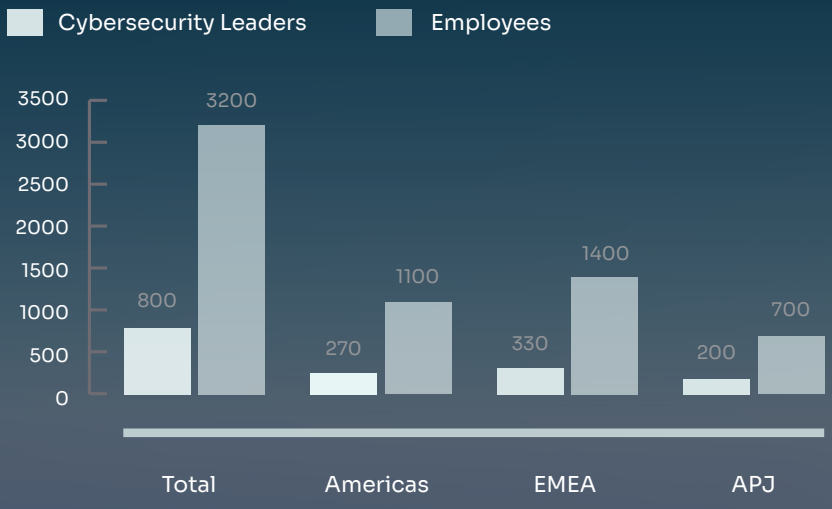
The workforce has changed. Employees and AI agents now operate as a single, interconnected layer of defense. That changes what good security looks like. Organizations that align awareness, behavior and culture around a clear picture of their risk do more than simply managing exposure. They create hard targets. The threats won’t disappear. They’ll run into better prepared organizations.

Get In Touch

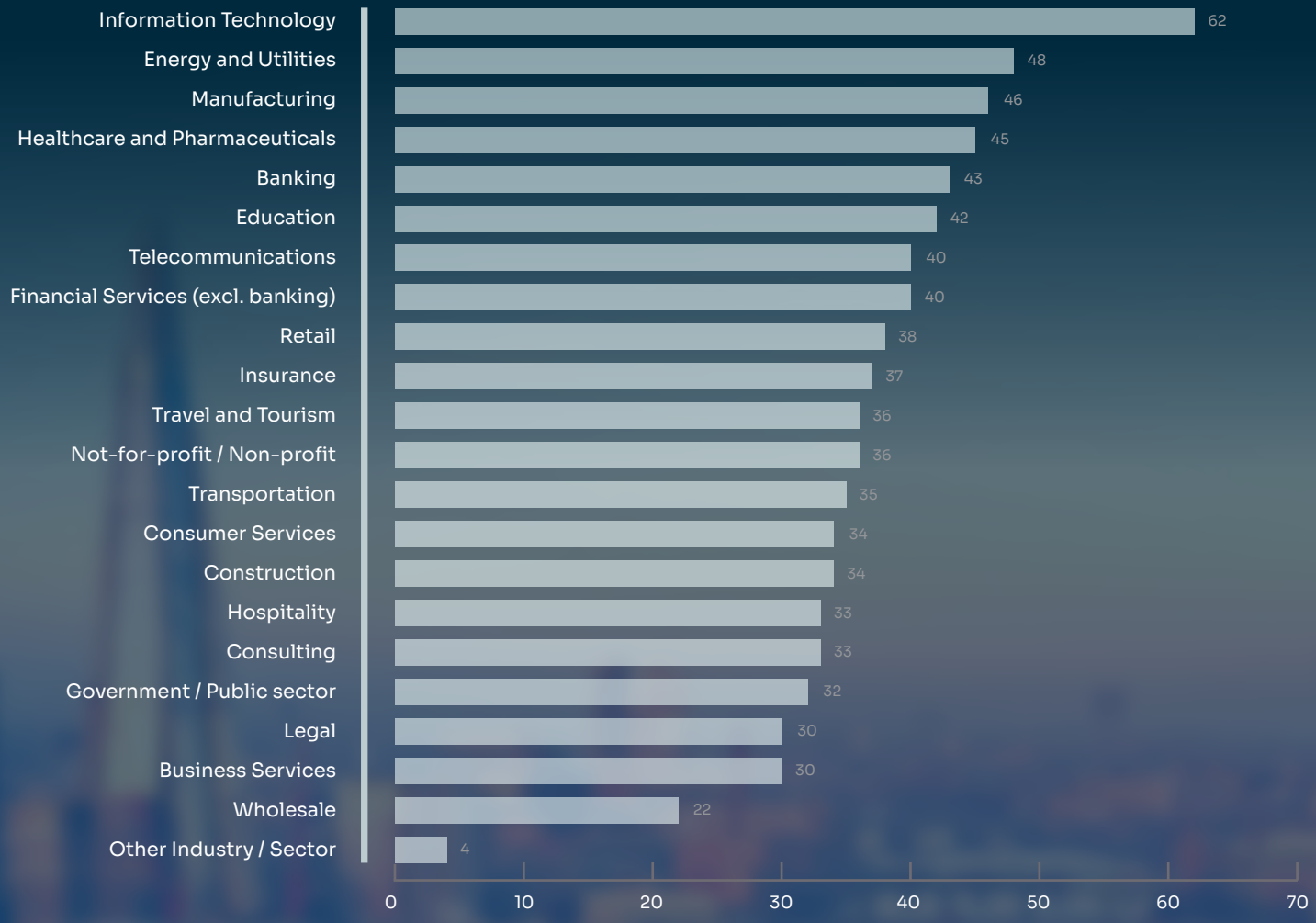
Methodology

This research is based on a global survey of 4,000 professionals, including 800 security decision makers and 3,200 employees, across the Americas, EMEA and APJ regions. Respondents represent organizations with 250 or more employees and span both private and public sectors among a wide range of industries such as information technology, healthcare, consumer services and others.

Same Size Breakdown: Regions



Sample breakdown: Industry segment (Cybersecurity Leaders, Total: 800)



Sample breakdown: Industry segment (Employees, Total: 3200)

About KnowBe4

KnowBe4 empowers the modern workforce to make smarter security decisions every day. Trusted by more than 70,000 organizations worldwide, KnowBe4 is the pioneer of digital workforce security, securing both AI agents and humans. The KnowBe4 Platform provides attack simulation and training, collaboration security, and agent security powered by AIDA (Artificial Intelligence Defense Agents) and a proprietary Risk Score. The platform leverages 15-years of behavioral data to combat advanced threats including social engineering, prompt injection, and shadow AI. By securing humans and agents, KnowBe4 leads the industry in workforce trust and defense.

More information at www.KnowBe4.com.

Follow us on Linkedin

Follow us on X

About Vanson Bourne

Vanson Bourne is an independent specialist in market research for the technology sector.
Their reputation for robust and credible research-based analysis is founded upon rigorous research principles and their ability to seek the opinions of senior decision makers across technical and business functions, in all business sectors and all major markets.

For more information, visit www.vansonbourne.com.

knowbe4

KnowBe4, Inc. | 33 N Garden Ave, Suite 1200, Clearwater, FL 33755
855-KNOWBE4 (566-9234) | www.KnowBe4.com | Sales@KnowBe4.com