Microsoft Security

Accelerate AI transformation with strong security

The path to securely embracing AI adoption in your organization

Foreword 3

Introduction 4

The push for rapid AI adoption 6

Facing the new realities of security for AI 10

Amplified risks of data leakage in AI systems 13

Increased threats with rapid AI development 14

Emerging risks and new challenges 15

The security transformation: A path to a secure yes to AI 16

4 steps to implementing effective security for AI 18

Step 1: Form a dedicated security team for AI 19

Step 2: Optimize resources to secure GenAI 22

Step 3: Implement a Zero Trust strategy 23

Step 4: Adopt a comprehensive security solution for AI 25

The path forward 27

Appendix 29

Report Foreword by Michal Braverman-Blumenstyk, Corporate Vice President, Microsoft Security Division CTO, Israel R&D Center Managing Director

Since the emergence of generative AI (GenAI), more and more companies have been adopting this powerful new technology and realizing its potential to transform their businesses. According to Microsoft's Work Trend Index, the use of GenAI has nearly doubled in the last six months, with 75% of global knowledge workers using it. Organizations are using GenAI not only to boost productivity, increase revenue, and reduce costs, but also to drive innovation. At Microsoft, we're seeing organizations use Microsoft Copilot to innovate in a variety of ways—they're accelerating rare disease research, maximizing value for clients, powering better patient care, and more.

GenAI is already driving significant advancements and transformation within organizations, but the rapid pace of adoption of this new technology comes with a host of new security concerns. And while security and risk leaders want to be able to say "yes" to their company's efforts to innovate, they want to be sure that the right security measures and solutions are in place.

At Microsoft, we wholeheartedly agree with a security first approach as we prioritize security above all else. That's why we're focused on helping customers use and build trustworthy AI that is secure, safe and private. It's also why we want to provide information that can be used to help make crucial decisions about how to secure and govern AI in organizations. Through conversations with many security and risk leaders about their excitement and concerns regarding GenAI, I realized that offering industry insights and data points can help clarify the path forward to confident AI adoption.

Microsoft Security, in partnership with MDC Research Group, conducted a study to learn how organizations approach AI adoption, use, development, and security. Based on survey responses from over 400 enterprise IT and data security-decision makers, this report serves three main purposes. First, it provides security and risk leaders with an opportunity to see what their peers around the world are thinking about and doing in relation to AI use, adoption, and security. It answers questions such as: How widespread is the use of GenAI in organizations? How prevalent is GenAI app development?

Second, readers can discover which GenAI security risks are most on the minds of security and risk leaders, and learn more about the new and amplified risks that come with AI use and development. Perhaps most importantly, the report outlines some of the steps leaders are taking or planning to take to address these risks.

I hope this report helps you feel more confident as you move forward on the path to securing and governing AI, enabling your organization to say "yes" to AI innovation with a safe and secure foundation.

Michal Braverman-Blumenstyk

The balancing act between innovation and security

The adoption of generative AI (GenAI) is accelerating across industries. Organizations of all sizes are in a sprint to incorporate AI into their operations to improve productivity and business processes, and increase revenue. Although there is a sense of hope and excitement surrounding this revolutionary technology, many security and risk leaders find themselves attempting to strike a balance between two powerful forces at play.

On one side, business leaders are eager to adopt AI applications so they can drive innovation.

On the other side, security teams are grappling with the obstacle of strengthening defenses against the new and complex security challenges posed by AI.

Questions that keep security and risk leaders up at night include:

- What if the AI application inadvertently leaks sensitive data?
- What if it hallucinates and produces inaccurate information?
- What if it produces malicious content that could damage our reputation?

The secure "yes"

Security and risk leaders want to be able to say "yes" to their company's appeals to innovate with AI, but they want it to be a safe and secure "yes." They want to ensure that the right security measures, protocols, and solutions are in place, and they want to be sure that their valid concerns about data security, inaccurate outputs, harmful content, and more are addressed.

To explore the topics of AI adoption, use, development, and security in more detail, Microsoft Security conducted quantitative and qualitative research that included over 400 security and IT decision makers as well as a series of in-depth interviews with decision makers. While the research found that companies are rapidly adopting and developing their own AI applications, security leaders clearly have concerns about the new and amplified risks that using and developing AI applications entails.

However, while AI risks are evolving, so are security solutions and best practices for AI. This white paper includes several recommended steps to more effectively address security concerns about AI, many of which emerged from the survey findings. More details about those will be provided later. But the first step is to examine the current AI landscape.

Microsoft Security

01

The push for rapid AI adoption

The push for rapid AI adoption

C-suite executives in companies around the world are pushing for the rapid adoption of GenAI. These leaders view GenAI as a critical driver of innovation and know that it holds the potential to revolutionize various sectors and lead to significant advancements.

GenAI can enable faster and more accurate cancer diagnoses, offer personalized learning experiences tailored to individual student needs, detect fraudulent transactions in real time, optimize production processes, enhance customer service with GenAI-powered virtual assistants, and optimize energy consumption in smart grids. These are just a few examples of the transformative possibilities that GenAI can offer.

The potential for GenAI innovation is high, and so are adoption rates. Of the survey respondents who passed the original screening criteria, only 5% of respondents said that they are neither using nor developing GenAI and have no plans to do so. Conversely, 95% of respondents reported that they were either planning to or already use and/or develop GenAI.

GenAI user and/or developer
95%
Not GenAI user or developer
5%

A majority of companies are using and developing GenAI-related apps

Although the widespread adoption and use of GenAI applications was expected, one notable survey finding was the number of respondents who said their companies were both using and developing GenAI apps or planning to.

66% of respondents said that their organizations are not only using GenAI applications but are also planning to develop or are already developing their own GenAI apps.

66% **of organizations are developing or planning to develop their own Gen AI applications**

Over a quarter (26%) of respondents were actively using third-party SaaS or enterprise-ready GenAI applications and had implemented GenAI apps they had developed or customized. Another 28% were in the process of actively testing third-party SaaS or enterprise-ready GenAI applications, and they were actively in the process of developing or customizing GenAI apps.

Meanwhile, over three quarters (76%) of those currently in the "use only" category said they have preliminary plans to develop or customize in the next year.

Why companies are developing several GenAI apps

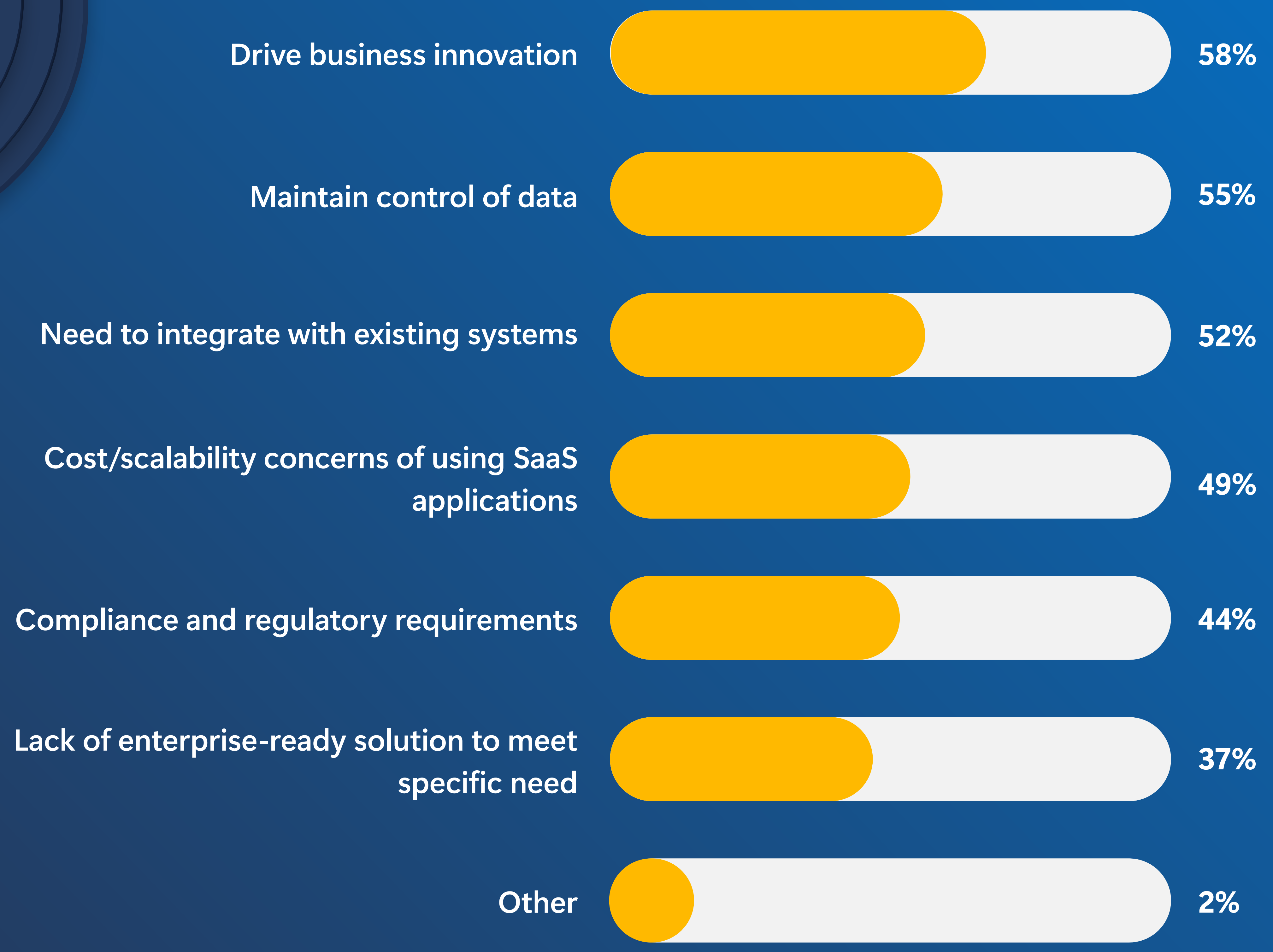
13.9

Average number of apps in development

Another notable survey finding was that companies that develop apps aren’t working on just one or two. Among companies who said they are developing or have developed apps, **the average number they were working on or had worked on was 13.9 apps.**

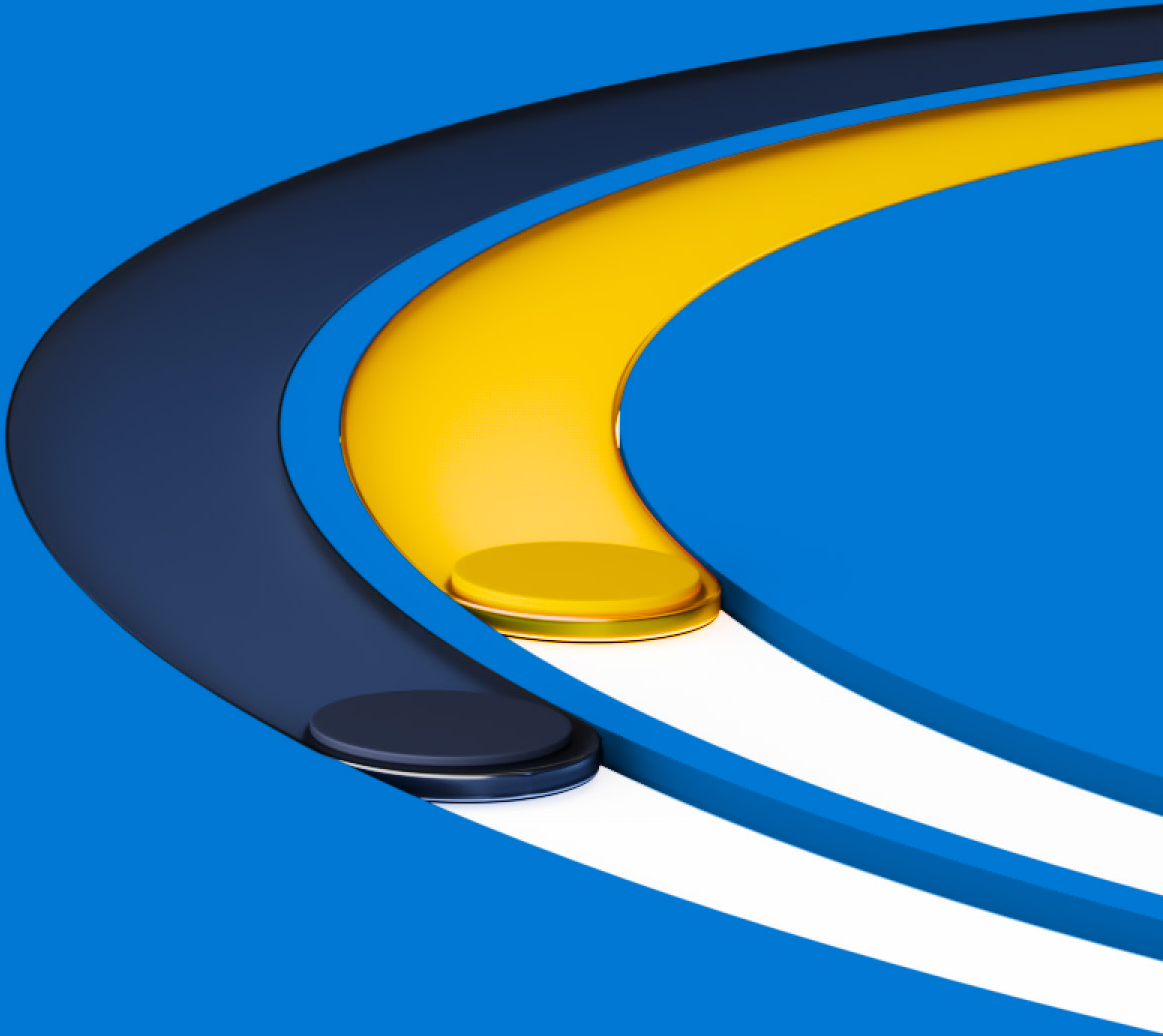
The top reason companies choose to develop or customize customer-facing GenAI applications is to **drive business innovation** (58%). Other reasons include wanting to maintain control of data (55%), needing to integrate with existing systems (52%), cost and scalability concerns (49%), and compliance and regulatory requirements (44%).

This fast-paced drive to develop new innovative applications can be exciting and potentially profitable for organizations, but it comes with new security concerns. With companies developing and deploying so many applications, AI components like orchestrators, models, and plug-ins can increase exposure and vulnerabilities for organizations. With all these elements in play, understanding new and amplified security concerns about AI is essential.



02

Facing the new realities of security for AI



Facing the new realities of security for AI

Addressing the evolving threat landscape is crucial to enabling trustworthy AI. GenAI introduces new attack surfaces, such as prompts, responses, training data, retrieval-augmented generation data, and models, effectively changing the risk landscape. In addition to managing traditional threat vectors, security and risk leaders also need to address amplified risks such as data leakage and data oversharing, and new risks such as prompt injections, hallucinations, and model vulnerabilities.

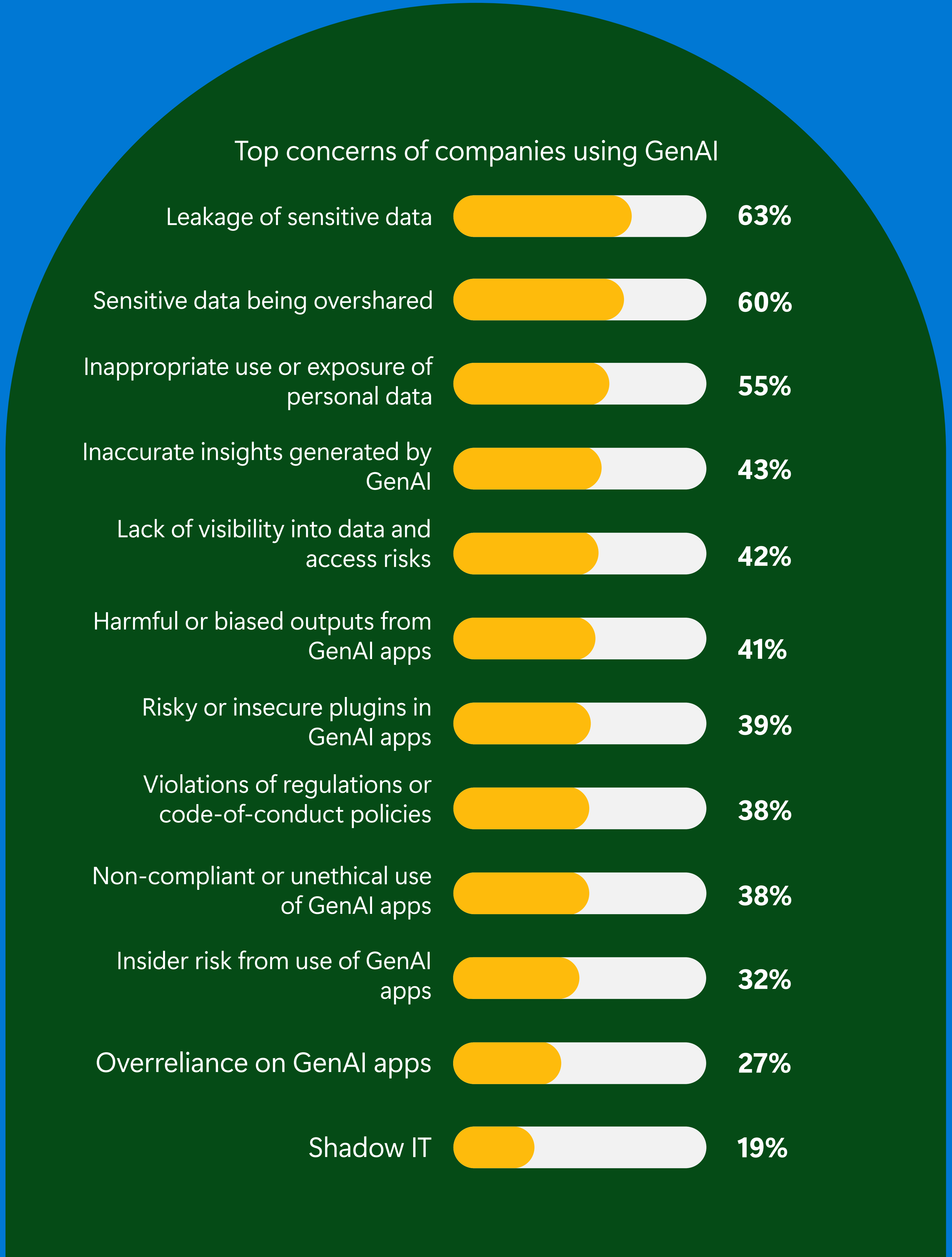
GenAI amplified and new risks
Indirect prompt injection
Jailbreak
Data leakage
Model vulnerability
Hallucinations
GenAI new attack surfaces
Prompts
Responses
AI orchestration
Training data
RAG data
Models
Plugins/skills
Your threat vectors
Application
Identity
Endpoints
Network
Data
Cloud

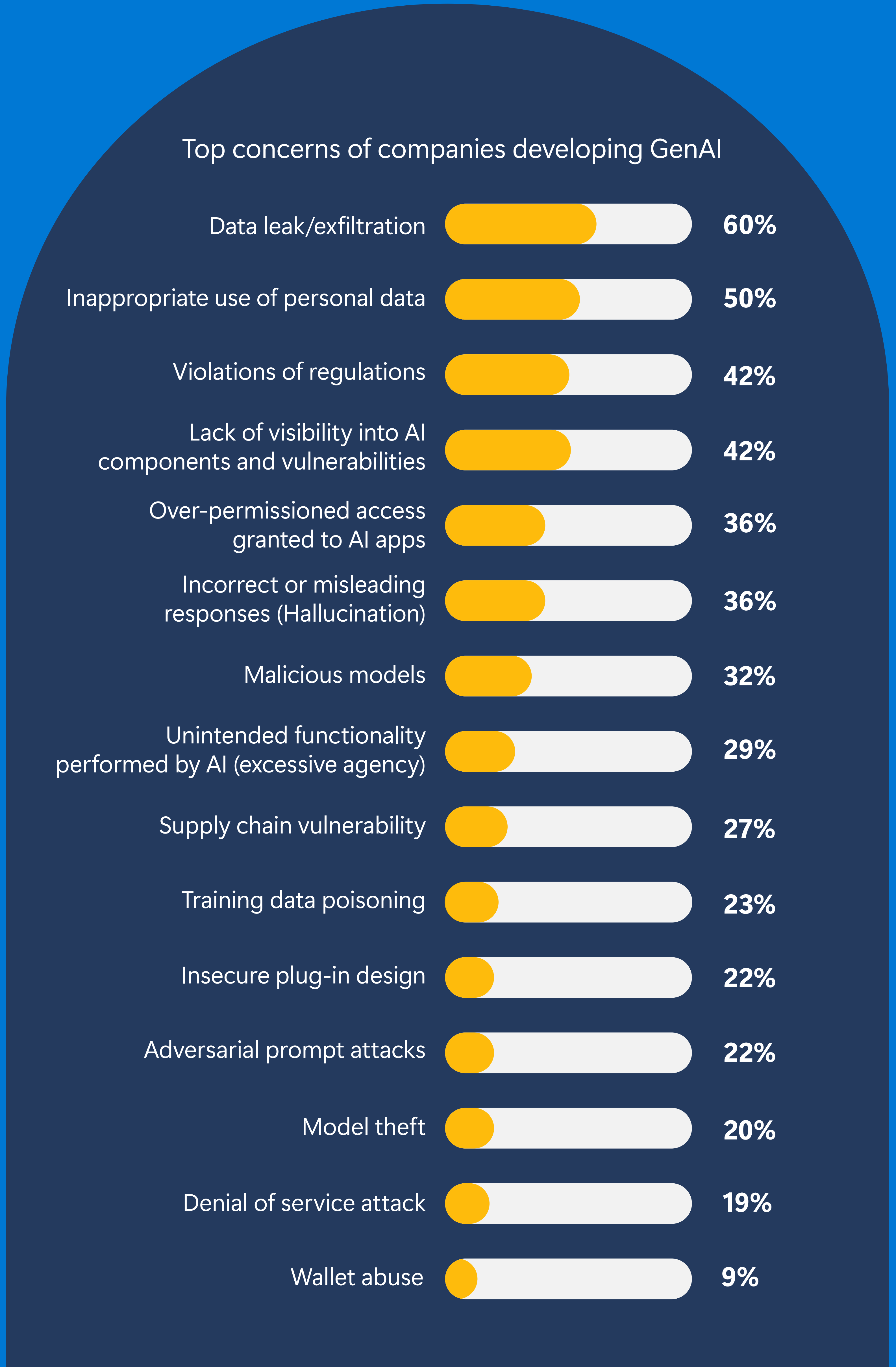
Security leaders' top concerns about GenAI

Security and risk leaders at companies using GenAI said their top concerns are data security issues, including leakage of sensitive data (63%), sensitive data being overshared, with users gaining access to data they're not authorized to view or edit (60%), and inappropriate use or exposure of personal data (55%). Other concerns include insight inaccuracy (43%) and harmful or biased outputs (41%).

In companies that are **developing or customizing GenAI apps**, security leaders' concerns were similar but slightly varied. Data leakage along with exfiltration (60%) and the inappropriate use of personal data (50%) were again top concerns. But other concerns emerged, including the violation of regulations (42%), lack of visibility into AI components and vulnerabilities (42%), and over-permissioned access granted to AI apps (36%).

Overall, these concerns can be divided into two categories: amplified and emerging security risks. The following sections examine these risks in more detail.





Amplified risks of data leakage in AI systems

As the volume of AI-generated content expands, the potential for data leakage and exposure increases as well. Organizations face heightened risks stemming from practices such as data oversharing and shadow IT.

Data oversharing and breaches: Data oversharing occurs when users inadvertently gain access to sensitive information through AI applications, often due to insufficient labeling policies or inadequate access controls. This might lead to unauthorized exposure of confidential data, posing significant risks to both individuals and organizations.

Without appropriate user training, the rapid proliferation of AI tools can also create environments in which users share or use data without fully understanding its sensitivity, compounding the risk of compliance violations and data breaches.

Shadow IT: With 78% of AI users bringing their own AI tools to work (BYOAI)[1], sometimes without the knowledge of the IT or security group within an organization, the risk of data leakage increases. When employees use third-party AI tools and paste sensitive information such as source code, meeting notes, and data spreadsheets into user prompts, they can inadvertently expose confidential company data outside of the company.

> We want to make sure that whatever data that we feed to it, it stays within [our company], and it’s not some proprietary information [that] gets leaked outside...
>
> Technical Decision Maker, IT

Increased threats with rapid AI development

As organizations attempt to keep up with the rapid evolution of AI technologies, the accelerated development and deployment of AI applications introduces elevated security risks for organizations.

Rushed deployments: Companies often face intense pressure to innovate quickly, which can result in inadequate testing, rushed deployments, and insufficient security vetting. This increase in the pace of development can leave critical vulnerabilities unaddressed, creating security risks once the AI system is in operation.

AI supply chain vulnerabilities: The AI supply chain is a complex ecosystem that presents potential vulnerabilities that could compromise the integrity and security of AI systems. Vulnerabilities in third-party libraries or models can expose AI systems to exploitation.

By "2028, open-source generative AI models will underpin more than 50% of enterprise GenAI use cases, up from less than 10% today."2 As organizations increasingly utilize open-source software to develop GenAI, components within the AI stack such as models and orchestrators can introduce vulnerabilities into their environments, which could be exploited by malicious actors.

AI misconfiguration: When developing and deploying AI applications, misconfigurations can expose organizations to direct risks, such as failing to implement identity governance for an AI resource, and indirect risks, such as vulnerabilities in an internet-exposed virtual machine, which could allow an attacker to gain access to an AI resource.

Emerging risks and new challenges

In addition to the amplification of existing security risks, AI can bring with it a host of emerging risks.

Hallucinations: An AI hallucination, in which an AI model generates false or misleading information, can pose risks to organizational integrity, and in high-stakes sectors like health care, finance, or legal services, they can lead to significant challenges. Hallucinations can also cause ethical and trust issues. Users must be able to trust that AI systems will provide accurate and reliable information, and hallucinations undermine this trust.

Harmful content: GenAI can generate offensive, dangerous, or legally non-compliant material. Malicious actors can use AI-produced deepfake video and audio content, fabricated news articles, and manipulated images to spread misinformation, sow discord, or harm reputations. The sophistication of AI models means they can produce highly realistic and convincing content, making the detection of such harmful outputs increasingly challenging.

Model theft: Model theft involves the illegal copying or theft of proprietary large language models (LLMs), which can erode competitive advantage and lead to financial losses as unauthorized parties can replicate models without incurring development costs. Brand reputation may also suffer from model theft if the stolen models are misused, and as language models grow more powerful, their theft also poses a significant security threat, such as unauthorized use and sensitive data exposure.

Prompt injections: In a prompt injection attack, a hacker disguises a malicious input as a legitimate prompt, causing unintended actions by an AI system. By crafting deceptive prompts, attackers can trick an AI model into generating outputs that include confidential information, making it challenging to detect and mitigate such threats. Users can deliberately exploit system vulnerabilities to elicit unauthorized behavior from the GenAI model or attempt to subvert safety and security filters. Direct injections overwrite system prompts, while indirect ones manipulate inputs from external sources.

Training data poisoning: Training data poisoning involves tampering with the data used to teach models to introduce vulnerabilities, biases, or backdoors. This can hurt a model's security and reliability and create risks like poor performance, system exploits, and harm to a company's reputation.

Excessive agency: Excessive agency allows an LLM-based system to perform harmful actions due to misinterpretations or unexpected errors in its decision-making. This vulnerability can compromise sensitive information, disrupt business operations, and result in security breaches, primarily when the model is granted too much decision-making power and autonomy.

Regulatory compliance: AI regulations like the European Union Artificial Intelligence Act (EU AI Act) are designed to ensure that AI systems are developed and used in a way that is safe, transparent, and responsible. Violations of the EU AI Act can cost companies as much as 35 million euros or 7% of annual turnover. This creates uncertainty for security and risk leaders as 62% of business leaders said they do not understand AI regulations that apply to their sector.[3]

These are a select few emerging security challenges. For more information about these and other AI risks, view a list of the top 10 risks for LLMs and GenAI Apps, compiled by the Open Worldwide Application Security Project (OWASP), and visit MITRE ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems).

03

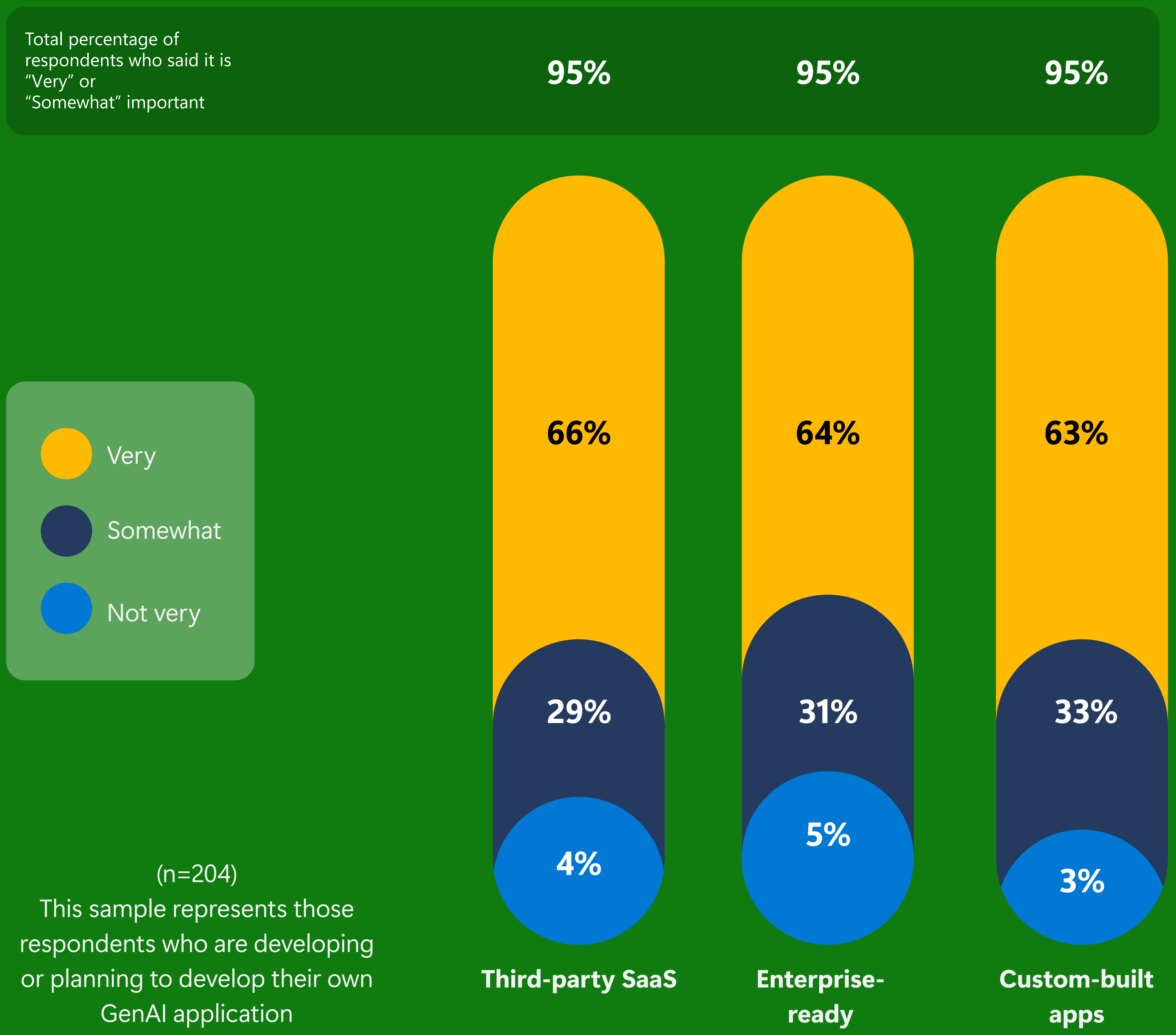
The security transformation: A path to a secure yes to AI

The security transformation: A path to a secure yes to AI

Security and risk leaders are clear they need to address these amplified and emerging threats as soon as possible. 95% of respondents agree that their company needs to have security measures in place for their AI apps in the next 12-24 months. They voiced a high concern for having security measures in place across all three AI application categories—third-party SaaS, enterprise-ready, and custom-built apps.

95% of security and risk leaders agree their company needs to have security measures in place for their AI apps including third-party SaaS, enterprise-ready, and custom-built apps in the next 12-24 months.

Level of urgency for having security measures in place for AI



(n=204)
This sample represents those respondents who are developing or planning to develop their own GenAI application

4 steps to implementing effective security for AI

As awareness of the risks associated with the rapid implementation of GenAI increases, many organizations are responding proactively by dedicating substantial resources to enhance their security measures. Security and risk leaders can take several actionable steps to create a path toward safe and secure AI innovation.

These recommended practices focus on fostering a collaborative environment and implementing effective security measures that will support GenAI advancements while safeguarding organizational interests.

01

Form a dedicated security team for AI.

02

Optimize resource allocation to secure GenAI.

03

Implement a Zero Trust strategy.

04

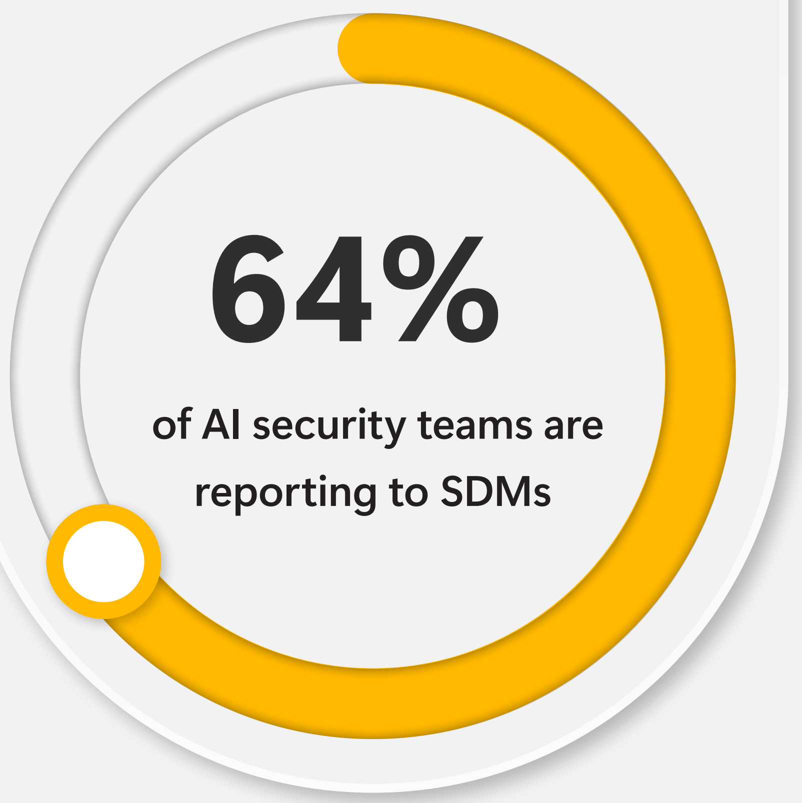
Adopt new dedicated security solutions for AI.

Step 1

Form a dedicated AI security team

A majority of companies recognize the need to form dedicated, cross-functional teams to manage the unique security challenges posed by AI. Dedicated security teams ensure that AI systems are rigorously tested, vulnerabilities are swiftly identified and mitigated, and security protocols are continuously updated to keep pace with evolving threats.

Eighty percent of survey respondents either currently have (45%) or plan to have (35%) a dedicated team to address security for AI. Over six in 10 said their teams will report to a security decision-maker, ensuring not only vigilant oversight but also strategic vision and leadership in addressing AI-related risks.



64%

of AI security teams are reporting to SDMs

80% of organizations currently have a dedicated team or plan to have to address security for GenAI.

80%

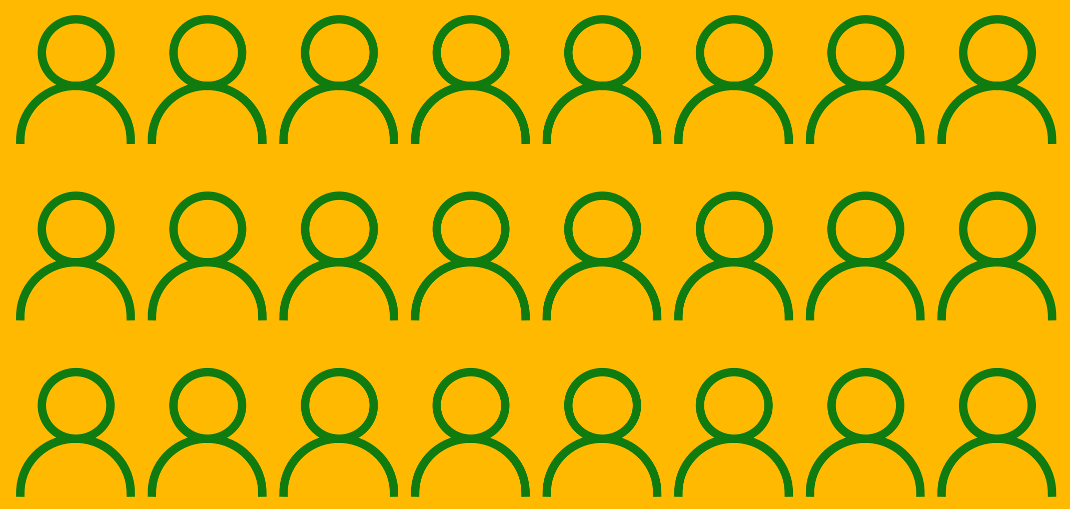
45%

of these organizations currently have a security team for AI.

35%

of these organization plan to have a security team for AI.

Notably, the median team size, or intended team size, of these dedicated security teams was 24 employees—underscoring the substantial resources that companies are committing to safeguarding their AI initiatives. When the size of company was factored in, team sizes varied.



Median team size

24

"I think [budget-wise], it's going to always sit in security...Your AI team is looking at the use cases and the data input and output and how things are shared, but security is not their focus, so it's really the security team that's going to own this and want to make sure that you're protecting things properly."

Security Decision Maker, Healthcare

Best practices for building a security team for AI

Here are a few best practices organizations can use to successfully build an effective cross-functional security team for AI.

01

Form an AI committee to foster collaboration across departments

Security for AI is a collective effort that goes beyond the IT department. Encourage collaboration among teams like security, IT, legal, compliance, and risk management to create comprehensive security strategies. Having varying perspectives and expertise will enhance the effectiveness of security protocols.

02

Hire diverse skill sets

Forming a successful security team for AI requires a balance of skills. Look for team members with expertise in data science, cybersecurity, software engineering, and machine learning. This diversity ensures that various aspects of security are covered, from technical development to threat prevention.

03

Establish clear roles and responsibilities

For effective productivity, clearly define each team member's role. Ensure everyone understands their specific responsibilities, which promotes accountability and avoids overlap in efforts.

04

Invest in continuous training and development

The rapid evolution of AI technologies mandates ongoing education for security teams. Provide access to training programs and workshops that focus on practices, emerging threats, and ethical considerations related to security for AI. This investment not only empowers team members but also ensures that the organization stays ahead of potential vulnerabilities.

Step 2

Optimize resources to secure GenAI

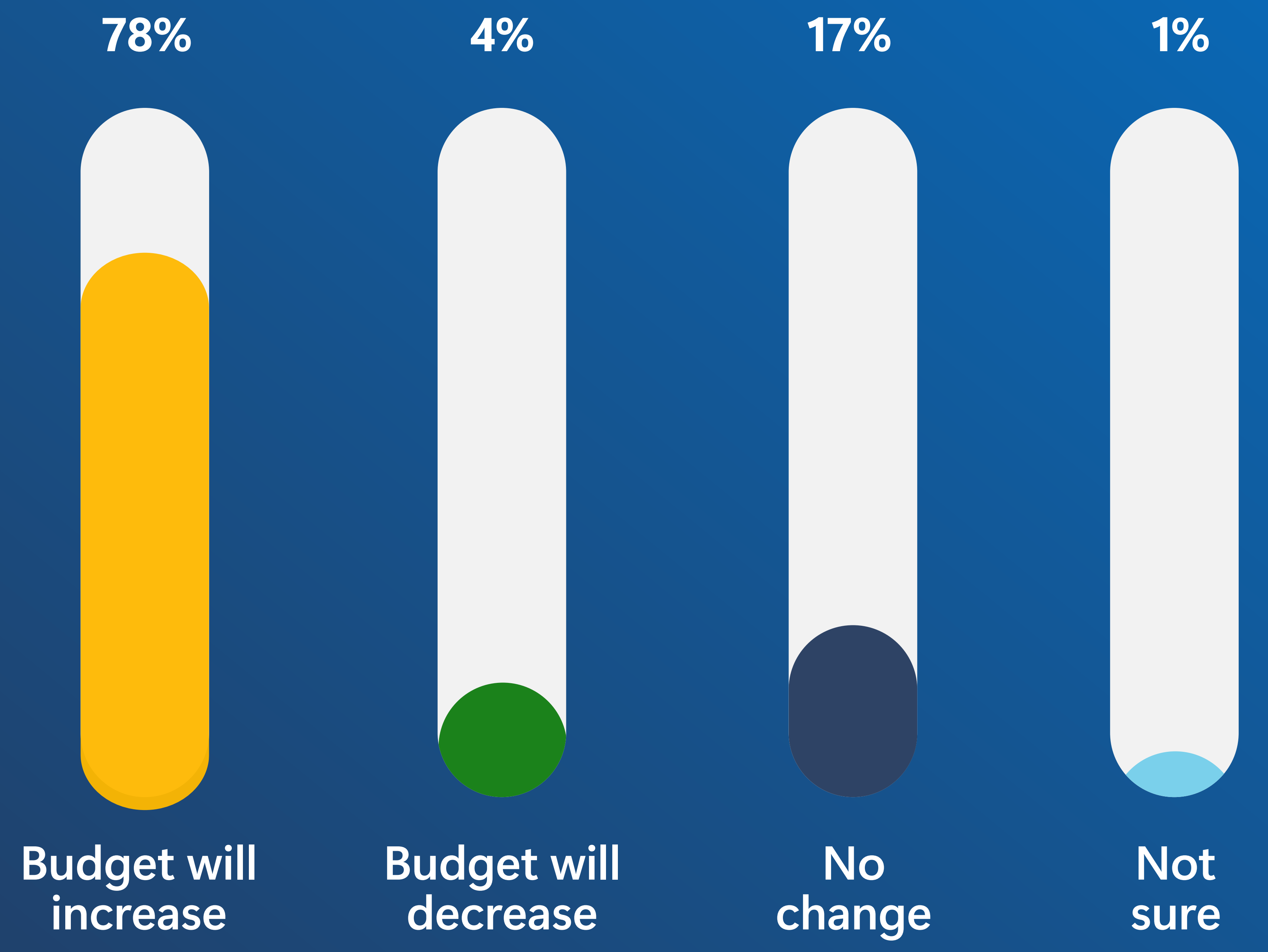
The introduction of AI applications within organizations is not only revolutionizing operations but also necessitating significant changes in resource and budget allocation, especially in IT security.

A significant majority of security and risk leaders (78%) believe their IT security budget will increase to accommodate the unique challenges and opportunities brought about by AI. This adjustment is crucial for several reasons. AI systems require a robust security infrastructure to operate securely. This might involve upgrading existing security systems, implementing more stringent access controls, and enhancing data security and governance. Additional resources might also be needed to meet emerging new AI regulatory requirements.

Allocating funds for compliance assessments, legal consultations, and audits becomes essential to align an organization's AI strategy to an industry framework and enable more secure, safe, and compliant AI usage and systems. Prioritizing funds for ongoing employee training and skills development—which could include specialized training on security tools for AI, risk management strategies, and ethical considerations in AI use—is also important to consider when allocating budget and resources.

78%
believe their IT security budget will increase with the adoption of GenAI applications.

Security for AI - budget expectations



Step 3

Implement a Zero Trust strategy

When preparing for AI adoption, a Zero Trust strategy provides security and risk leaders with a set of principles that help address some of their top concerns, including data oversharing or overexposure and shadow IT. A Zero Trust approach shifts from a network-centric focus to an asset and data-centric focus and treats every access request as a potential threat, regardless of its origin.

Zero Trust constantly validates the identities of every user and device, ensuring that only those with clear permissions can reach sensitive information. By dynamically adjusting security measures based on real-time assessments, Zero Trust minimizes the risk of data leakage and protects an organization from both internal and external threats. Continuous verification, least privilege access, and dynamic risk management are the cornerstones of this approach, providing a robust and adaptable security framework that supports the success of an organization's end-to-end security.

By embracing Zero Trust, organizations can secure their AI deployments and know that their security is continuously validated and protected. Zero Trust empowers organizations to embrace AI confidently, ensuring that AI's powerful capabilities are harnessed safely and effectively.

Zero Trust principles

Verify explicitly

Diligently verify all identities accessing AI applications, and assess every application used, deployed, and developed to ensure security integrity. By defining and monitoring both intended and unintended activities, organizations can maintain a strong defense against unauthorized access.

Use least privilege access

Limit AI systems to only access data necessary for intended uses by authorized users and ensure that AI agents operate with the minimum privilege to perform intended tasks—typically with just-in-time and just-enough-access (JIT/JEA) and risk-based policies like adaptive access control.

Assume Breach

Breaches are inevitable, so this principle focuses on minimizing their impact. To proactively design effective controls to reduce risks, operate under the assumption that each AI prompt could have malicious intent, responses might inadvertently leak data, and that AI components may possess vulnerabilities.

Step 4

Adopt a comprehensive security solution for AI

As AI adoption continues to expand across industries, the need for comprehensive, dedicated security solutions has become increasingly apparent. AI introduces specific risks that traditional security measures might not fully address. Security for AI is designed to mitigate these risks.

A significant majority of companies plan to procure specialized tools and platforms to secure both the usage and development of AI applications.

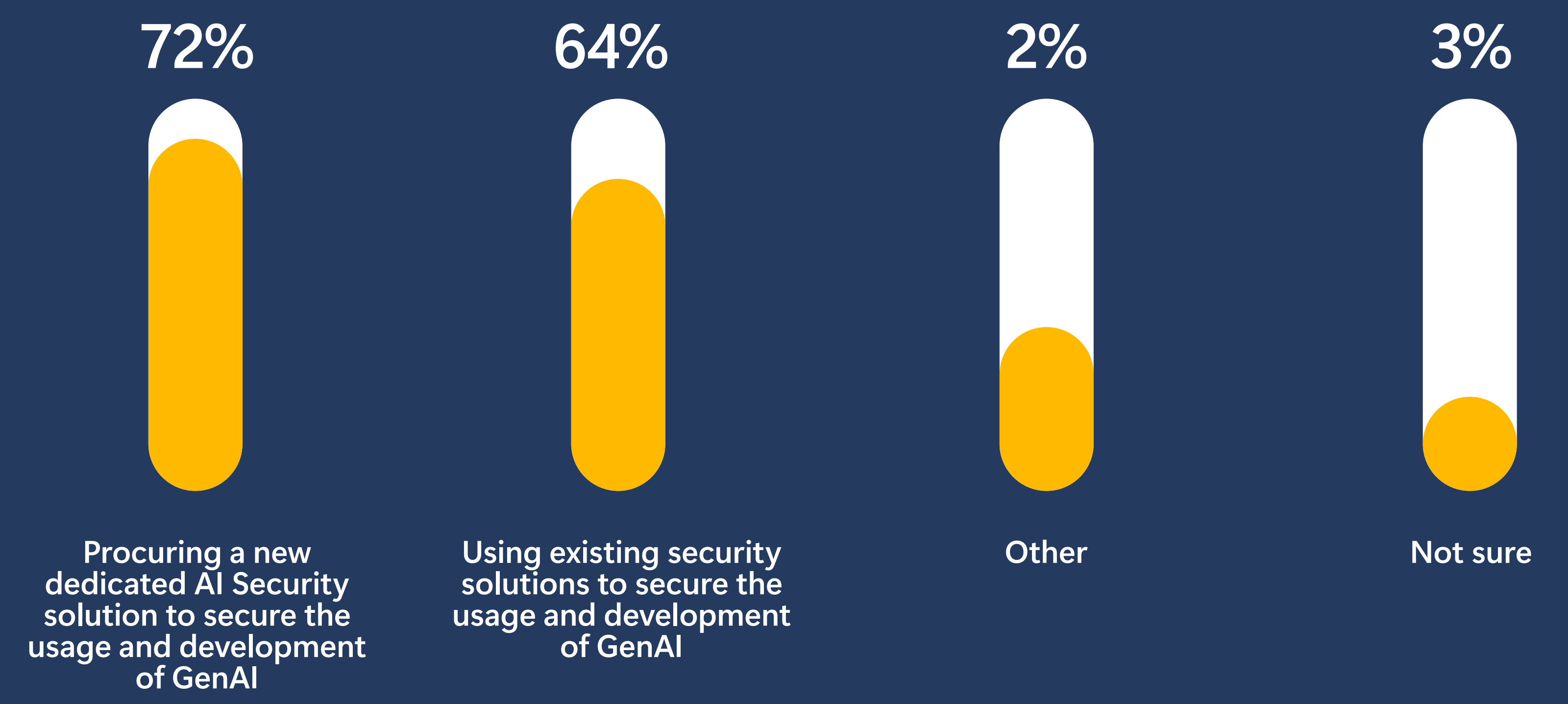
Organizations are looking for a comprehensive security solution set among new and existing solutions

When asked how they plan on securing and protecting the usage and development of AI applications in their organizations, a majority of survey respondents (72%) said they plan to procure a new dedicated security solution to secure the usage and development of AI, while 64% stated they plan to use existing security solutions to secure AI.

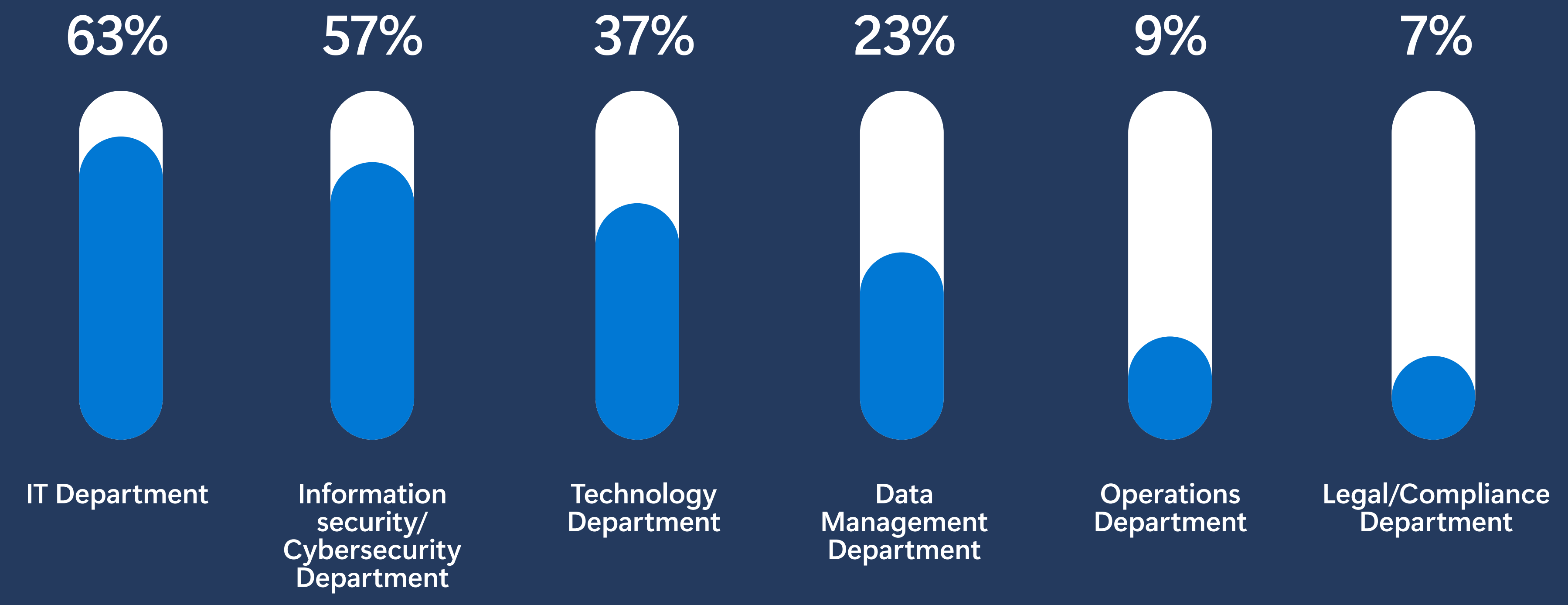
IT and security leaders believe that the primary budget contributors for new solutions for the protection and governance of AI will be IT departments (63%) and information security/ cybersecurity departments (57%).

These findings show that in addition to continuing to leverage existing security solutions, organizations see the need to look for new solutions that can help address the amplified and emerging risks of AI.

Security plans for AI

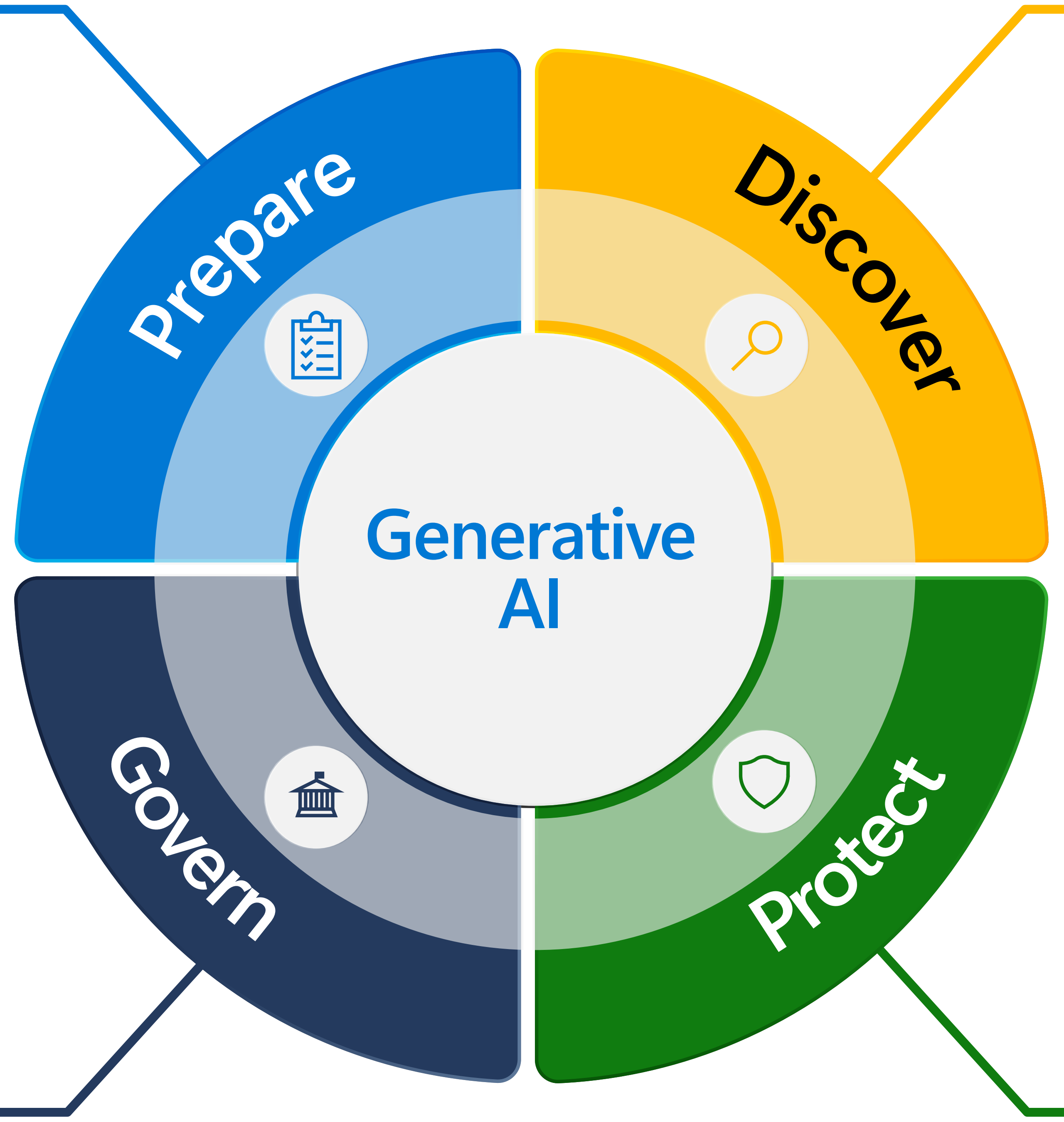


Security for AI - budget contributors



Key elements of a comprehensive security solution for AI

When evaluating security solutions for AI, security and risk leaders should consider comprehensive solutions that help companies prepare environments for secure adoption, discover AI-related risks, protect AI systems, and govern AI data to comply with regulations.



Organizations should choose a security solution for AI that helps prepare their environments for AI adoption with confidence. This includes classifying and labeling data within their environments, as well as ensuring robust identity and access governance to support a Zero Trust strategy.

Once AI adoption is underway, a security solution for AI should help security teams continuously discover security, safety, and privacy risks–allowing them to proactively design and adjust controls and policies to address evolving threats.

To comply with regulatory requirements, a security solution for AI should include guidance and assessments to evaluate, implement, and strengthen compliance controls, alongside enterprise-ready compliance solutions to help govern AI interactions more effectively.

A comprehensive solution is designed to continuously protect AI and data as developers deploy AI applications, and as users and customers interact with these applications. This includes protecting sensitive data in AI prompts and responses, and detecting, blocking, and responding to threats such as prompt injections.

The path forward

In a rapidly evolving technological landscape, security and risk leaders can strike a balance between their company's needs for innovation and security by taking proactive measures to successfully mitigate potential risks associated with AI technologies. This enables organizations to confidently embrace AI as a powerful tool for transformation and growth without compromising security.

To effectively tackle the complexities of security for AI, companies are forming security teams for AI, optimizing resources, and implementing a robust Zero Trust strategy that helps organizations continuously evaluate and respond to threats. They are also adopting comprehensive security solutions for AI that empower them to proactively prepare their environments, discover AI-specific risks, protect their AI systems, and govern AI to ensure compliance with ever-evolving regulations.

Through this strategic, multi-faceted approach, organizations can tap into the remarkable potential of GenAI and drive meaningful innovation without sacrificing the high standards of security required in today's interconnected world.

Learn more about how Microsoft can help secure and govern AI to accelerate your AI transformation with confidence.

Appendix

Methodology

Quantitative research for this study was conducted by Microsoft Security in partnership with MDC Research Group in June 2024. A total of 402 surveys were completed by enterprise business IT and data security-decision makers, with consumers from the US, UK, Canada, Australia, and other global English markets. Survey respondents were employed full time at an enterprise company with 1,000 or more employees.

In a separate study, qualitative and quantitative research was conducted by Microsoft Security in partnership with Answers Research in July 2024. A total of 15 60-minute in-depth interviews were conducted with participants who were employed full-time in an ITDM, SDM, or TDM role, as a C-suite, C-1, or C-2 level in a company with 5,000+ employees. A total of 409 surveys were completed by enterprise IT, technical, and data security-decision makers in the US in a company with 1,000+ employees.

References

1. 2024 Work Trend Index Annual Report, May 2024
2. Gartner®, Innovation Guide for Generative AI Models, Arun Chandrasekaran, Arnold Gao, Ben Yan, August 26, 2024. GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.
3. First Annual Generative AI study: Business Rewards vs. Security Risks, ISMG, Q3 2023