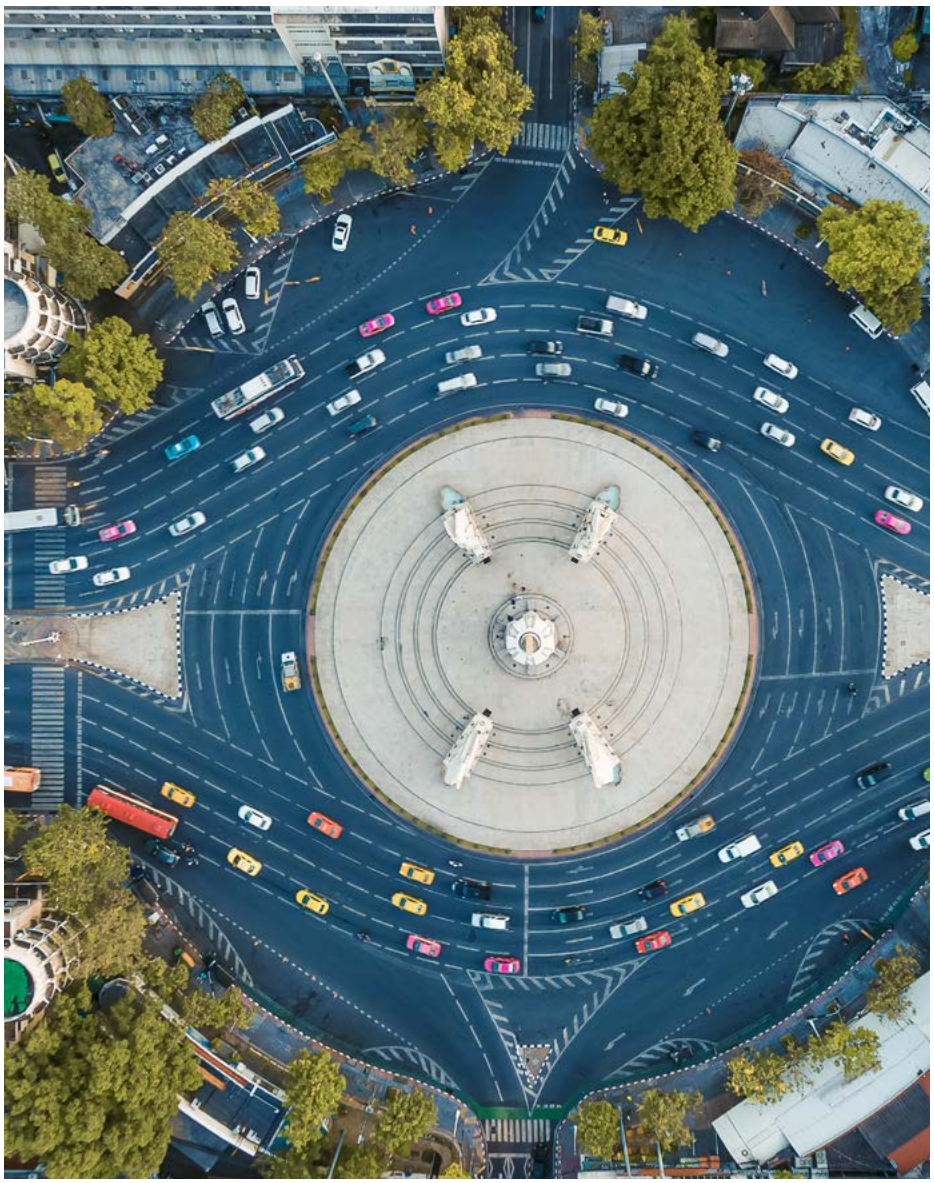


Securing generative AI

What matters now

IBM | aws

How AWS can help

For over 15 years, Amazon Web Services has been the world's most comprehensive and broadly adopted cloud offering. Today, we serve millions of customers, from the fastest growing startups to the largest enterprises, across a myriad of industries in practically every corner of the globe. We've had the opportunity to help these customers grow their businesses through digital transformation efforts enabled by the cloud. In doing so, we have worked closely with the C-suite, providing a unique vantage point to see the diverse ways executives approach digital transformation—the distinct thought processes across C-suite roles, their attitudes and priorities, obstacles to progress, and best practices that have resulted in the most success. For more information, please visit: https://aws.amazon.com

How IBM can help

IBM Security® works with you to help protect your business with an advanced and integrated portfolio of enterprise cybersecurity solutions and services infused with AI. Our modern approach to security strategy uses zero-trust principles to help you thrive in the face of uncertainty and cyberthreats. For more information, please visit: https://ibm.com/security

Key takeaways

Generative AI solutions can be as vulnerable as they are valuable if security is an afterthought.

- Only 24% of current generative AI projects are being secured.

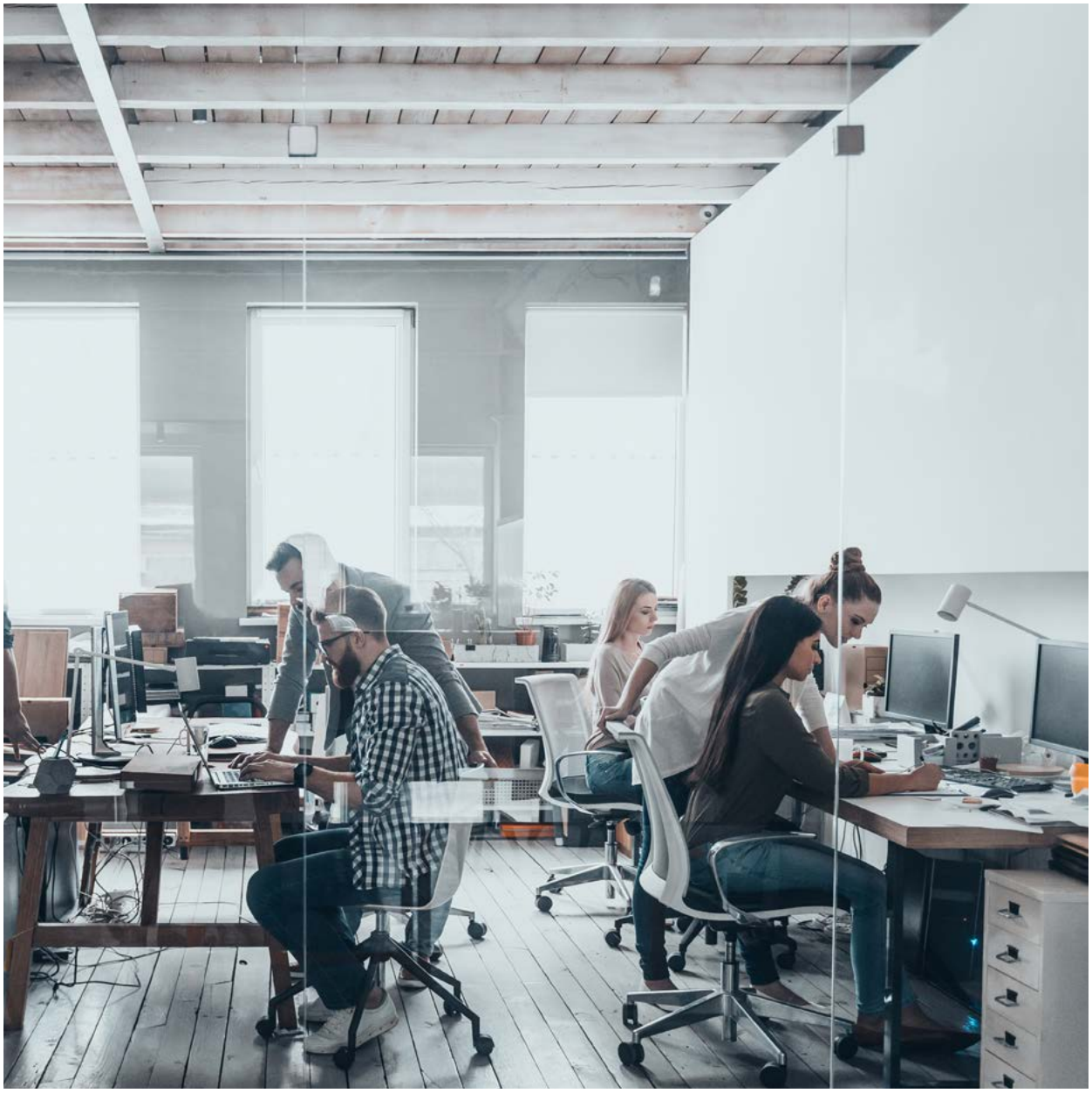
  While a majority of executives are concerned about unpredictable risks impacting gen AI initiatives, they are not prioritizing security.

- A changing threat landscape demands a new approach to securing AI.

  Built on a foundation of governance, risk, and compliance, securing AI infrastructure means securing applications, data, models, and model usage.

- Organizations are turning to third-party products and partners for over 90% of their gen AI security requirements.

  Just as with the transition to cloud, partners can help assess needs and manage security outcomes.



Innovation versus security: It's not a choice, it's a test

As organizations rush to create value from generative AI, many are speeding past a critical element: security. In a recent study of C-suite executives, the IBM Institute for Business Value (IBM IBV) found that only 24% of current gen AI projects have a component to secure the initiatives, even though 82% of respondents say secure and trustworthy AI is essential to the success of their business. In fact, nearly 70% say innovation takes precedence over security.

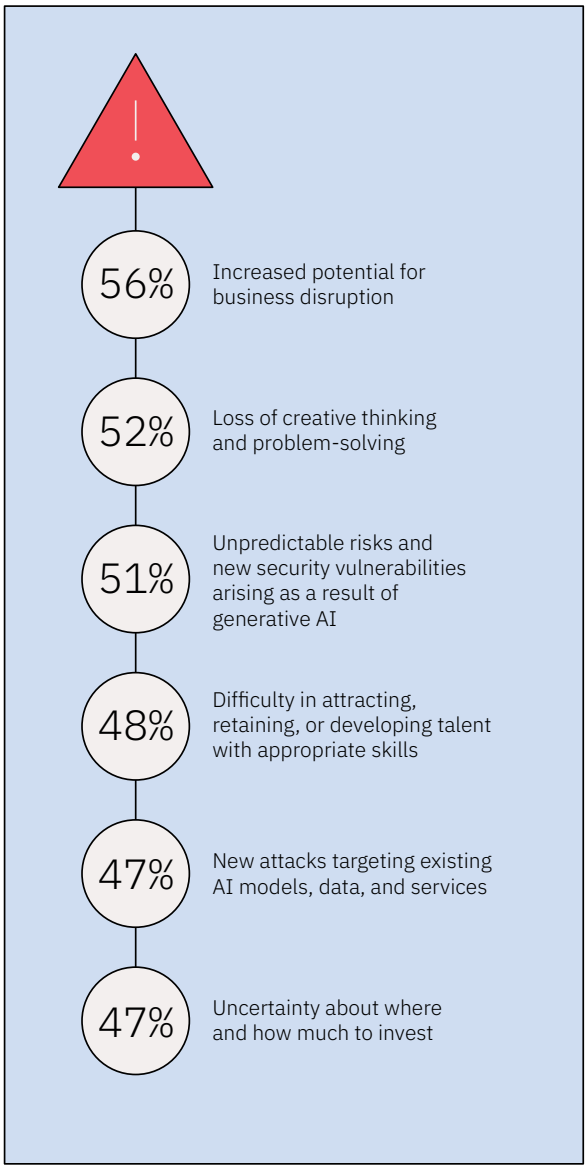
This perceived trade-off contrasts with executives' views of the wide-ranging risks of gen AI. Security vulnerabilities are among their biggest areas of concern (see Figure 1).

These worries are well-founded. Cybercriminals are already benefitting from both generative and traditional AI (see Perspective, "Understanding the generative AI threat landscape"). More realistic email phishing tactics and deepfake audios are making headlines, as are data leaks from employees' careless use of public tools such as ChatGPT.[1]

Looking ahead, potential threats to critical AI systems are even more troubling. As AI-powered solutions become more capable and more ubiquitous—integrated within critical infrastructure such as healthcare, utilities, telecommunications, and transportation—they could be as vulnerable as they are valuable, especially if security is an afterthought.

FIGURE 1

Executives expressed a broad spectrum of concerns regarding their adoption of gen AI.



Q. What are you most concerned about in adopting generative AI?

While a consolidated AI threat surface is only starting to form, IBM X-Force® researchers anticipate that once the industry landscape matures around common technologies and enablement models, threat actors will begin to target these AI systems more broadly.[2] Indeed, that convergence is well underway as the market is maturing rapidly, and leading providers are already emerging across hardware, software, and services.[3]

The gap between executives' angst and action underscores the need for cybersecurity and business leaders to commit to securing AI—now. With new IBM IBV research showing many organizations are still in the evaluation/pilot stages for most generative AI use cases such as information security (43%) and risk and compliance (46%), this is the time to get ahead of potential threats by prioritizing security from the start.[4]

To address the need for more specific guidance on where to begin, the IBM IBV and IBM Security have teamed with Amazon Web Services (AWS) experts to share leading practices and recommendations based on recent research insights. Part one of this report provides a framework for understanding the gen AI threat landscape. In part two, we discuss the three primary ways organizations are consuming gen AI and the related security considerations. Part three explores resource challenges and the role of partners. Part four offers an action guide of practical steps leaders can take to secure AI across their organizations.

With many organizations still evaluating and piloting generative AI solutions, now is the time to get ahead of new security threats.

Perspective

Understanding the generative AI threat landscape[5]

Generative AI introduces new potential threat vectors and new ways to mitigate them. While the technology lowers the bar even further for low-skill threat actors, helping them develop more sophisticated exploits, it also enhances defenders' capacity to move faster with greater efficiency and confidence.

Red team

Social engineering and fraud
Allows more targeted, convincing phishing messages on a mass scale

Data theft
Enables autonomous theft of sensitive data and intellectual property, and evasion of antivirus software through AI-enhanced malware

Identify theft and impersonation
Makes it easier to pass through online filters and enable illegal activities such as fraudulent account creation

AI jailbreaks
Removes the guardrails on gen AI chatbots, so they trick victims into giving away personal data or login credentials

Password cracking
Uses publicly available data to generate possible passwords

Vulnerability exploits
Application, data, model, or infrastructure vulnerabilities, such as misconfigurations, accidental disclosures, and policy/controls oversights

Blue team

Continuous regulatory compliance
Provides a real-time view into security and compliance posture and automates compliance tasks

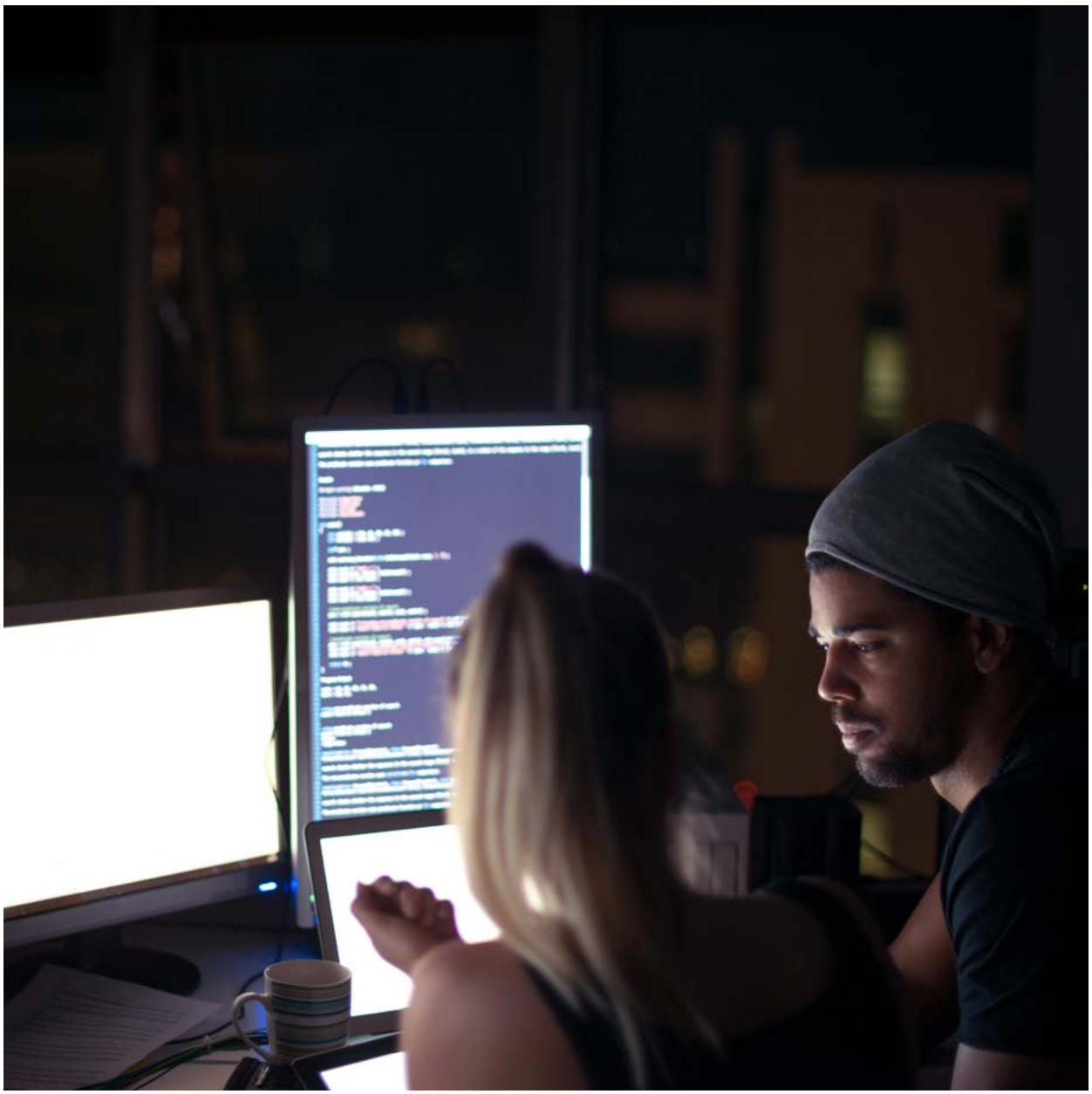
Case management
Generates summaries of security cases and incidents, and identifies similar cases for improved forensic analysis

Accelerated threat hunting
Detects threats based on natural language descriptions of cyber incident behaviors and patterns

Incident simulation and pen testing
Accelerates analysis of event inputs/outputs and generation of test scenarios

Data interpretation
Collates telemetry data across sources, and speeds analysts' understanding of security log data

API security
Transforms automation using API discovery, testing, and protection

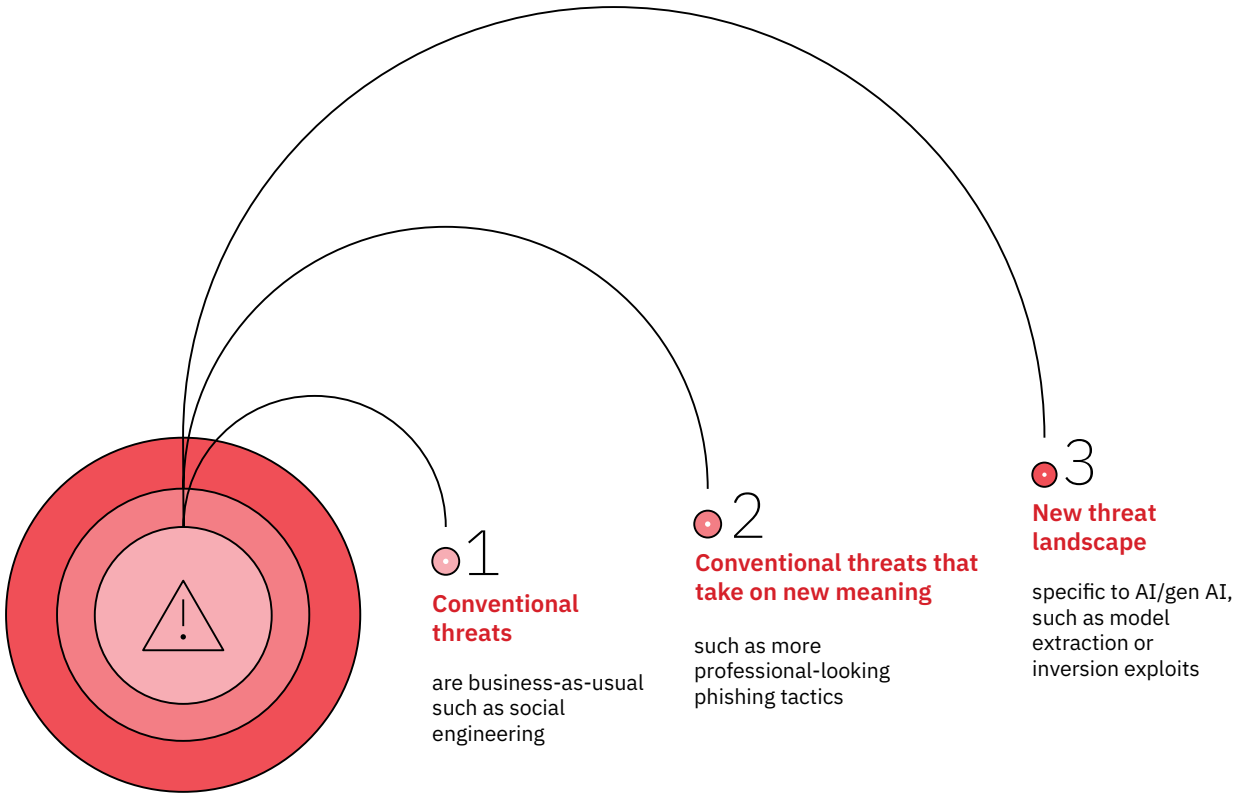


Seeing threats in a new light

For generative AI to deliver value, it must be secure in the traditional sense—in terms of the confidentiality, integrity, and availability of data.[6] But for gen AI to transform how organizations work—and how they enable and deliver value—model inputs and outputs must be reliable and trustworthy. While hallucinations, ethics, and bias often come to mind first when thinking of trusted AI, the AI pipeline faces a threat landscape that puts trust *itself* at risk. Each aspect of the pipeline—its applications, data, models, and usage—can be a target of threats—some familiar and some new (see Figure 2).[7]

FIGURE 2

Some emerging threats look familiar while others are entirely new.



Source: IBM Security.

Conventional threats, such as malware and social engineering, persist and require the same due diligence as always. For organizations that may have neglected their security fundamentals or whose security culture is still in the formative stages, these kinds of threats will continue to be a challenge.

Given the increasing adoption of AI and automation solutions by threat actors, organizations without a strong security foundation will also be ill-prepared to address the new twists on conventional threats introduced by gen AI. Take phishing emails as an example. With gen AI, cybercriminals can create far more effective, targeted scams—at scale.[8]

IBM Security teams have found gen AI capabilities facilitate upwards of a 99.5% reduction in the time needed to craft an effective phishing email.[9] This new breed of email threats should moderately impact companies with mature approaches to identity management, such as standard practices for least privilege and multifactor authentication as well as zero-trust architectures that restrict lateral movement. But those who lag in these areas run the risk of incidents with potentially devastating reach.[10]

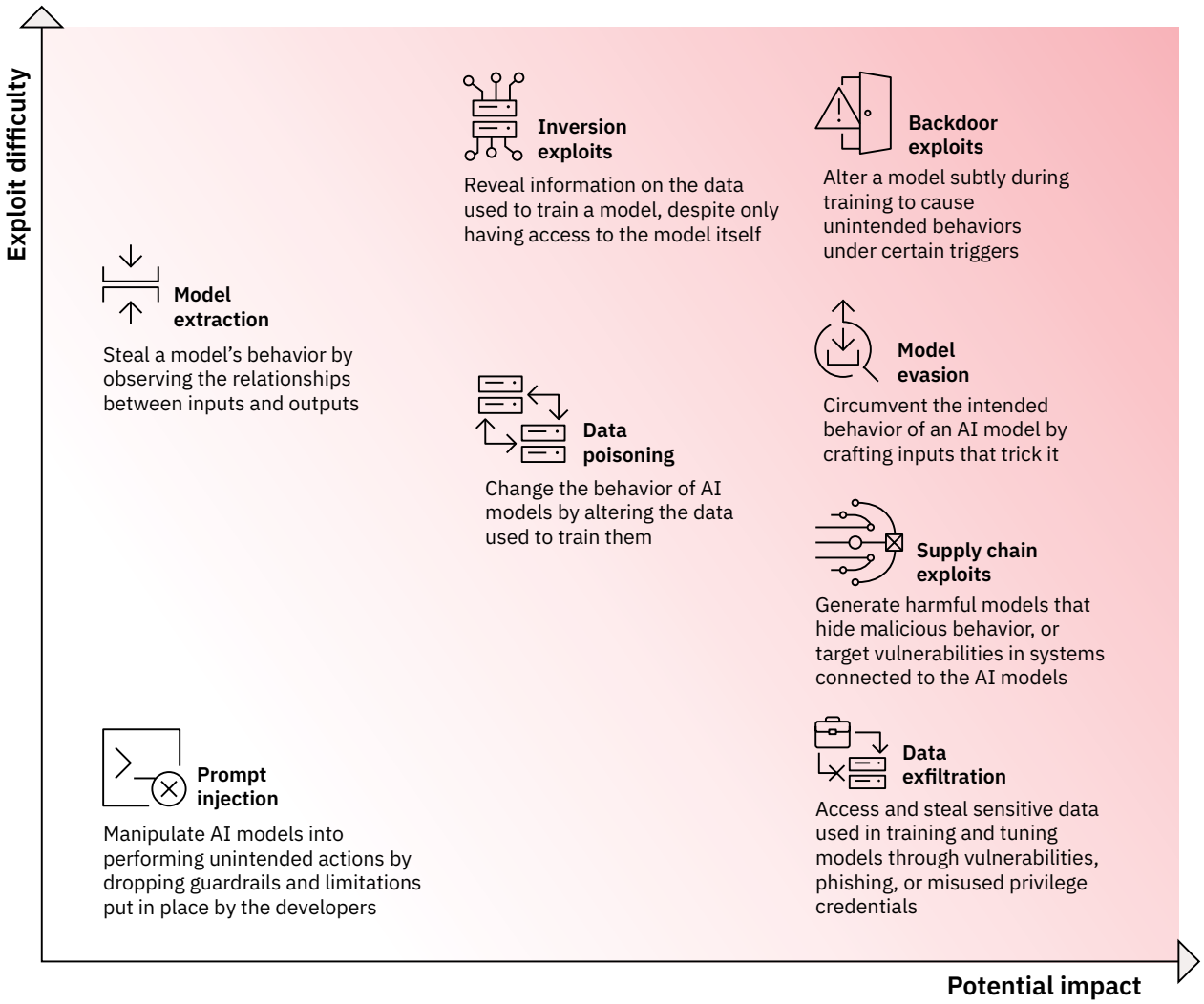
The reality is that security deficiencies are indeed impacting a significant number of organizations, as results from an IBM IBV survey of more than 2,300 executives suggest. Most respondents reported their organization's capabilities in zero trust (34%), security by design (42%), and DevSecOps (43%) are in the pilot stage.[11] These organizations will need to continue investing in core security capabilities as they are critical for protecting generative AI.

Organizations without a strong security foundation will also be ill-prepared to address the new twists on conventional threats introduced by gen AI.

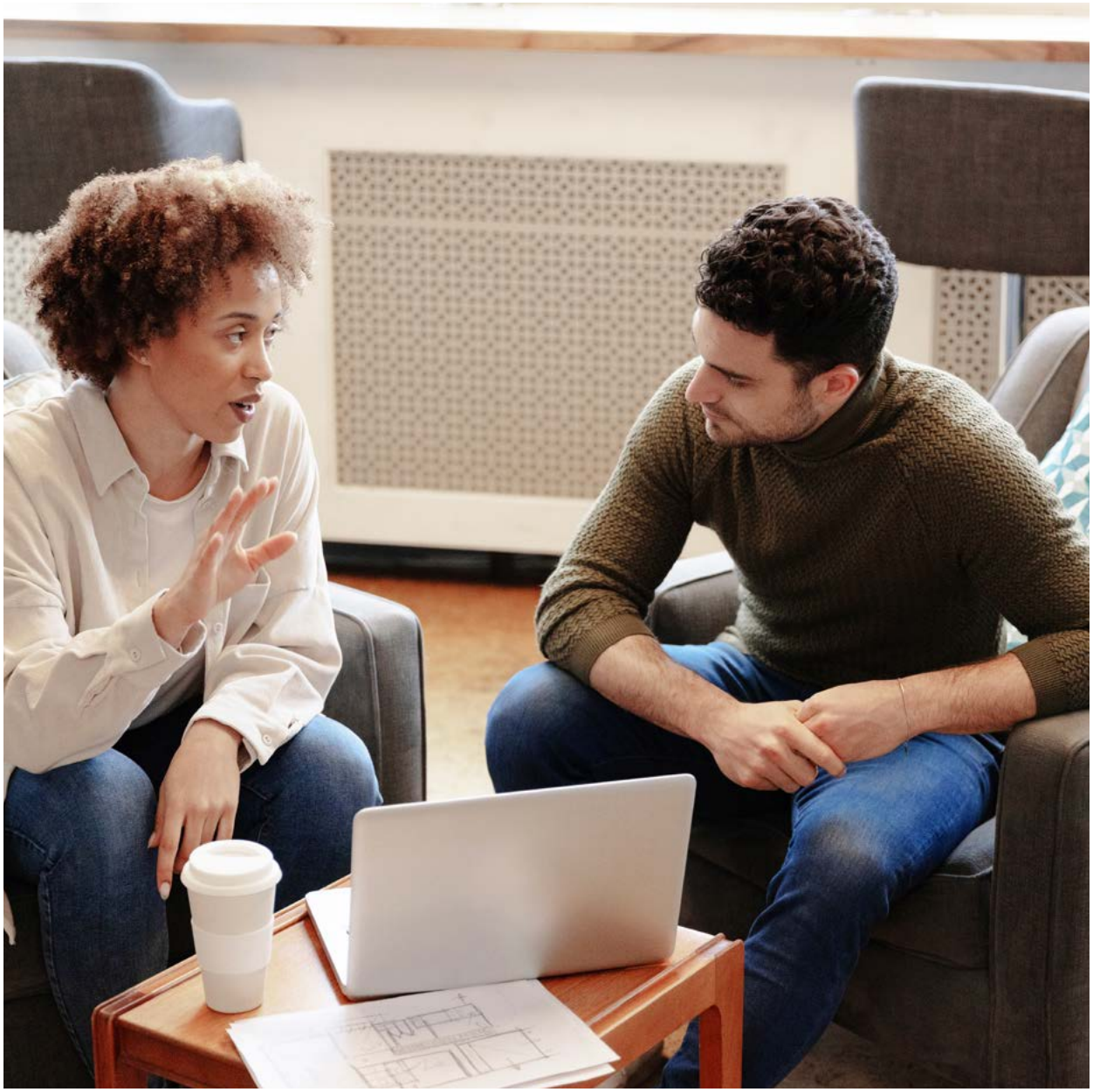
Lastly, a set of fundamentally new threats to organizations' gen AI initiatives is also emerging—a fact recognized by nearly half (47%) of respondents in our survey (see Figure 3). Prompt injection, for instance, refers to manipulating AI models to take unintended actions; inversion exploits cull information about the data used to train a model. These techniques are not yet widespread but will proliferate as adversaries become more familiar with the hardware, software, and services supporting gen AI.[12] As organizations move forward with gen AI solutions, they need to update their risk and governance models and incident response procedures to reflect these emerging threats. In a recent AWS Executive Insights podcast, security subject-matter experts emphasized that threat actors will go after low-hanging fruit first—threats with the greatest impact for the least amount of effort.[13] When choosing security investments, leaders should prioritize those use cases, such as supply chain exploits and data exfiltration.

FIGURE 3

Emergent threats to AI operations require updates to organizations' risk and governance models.



Source: IBM Security.

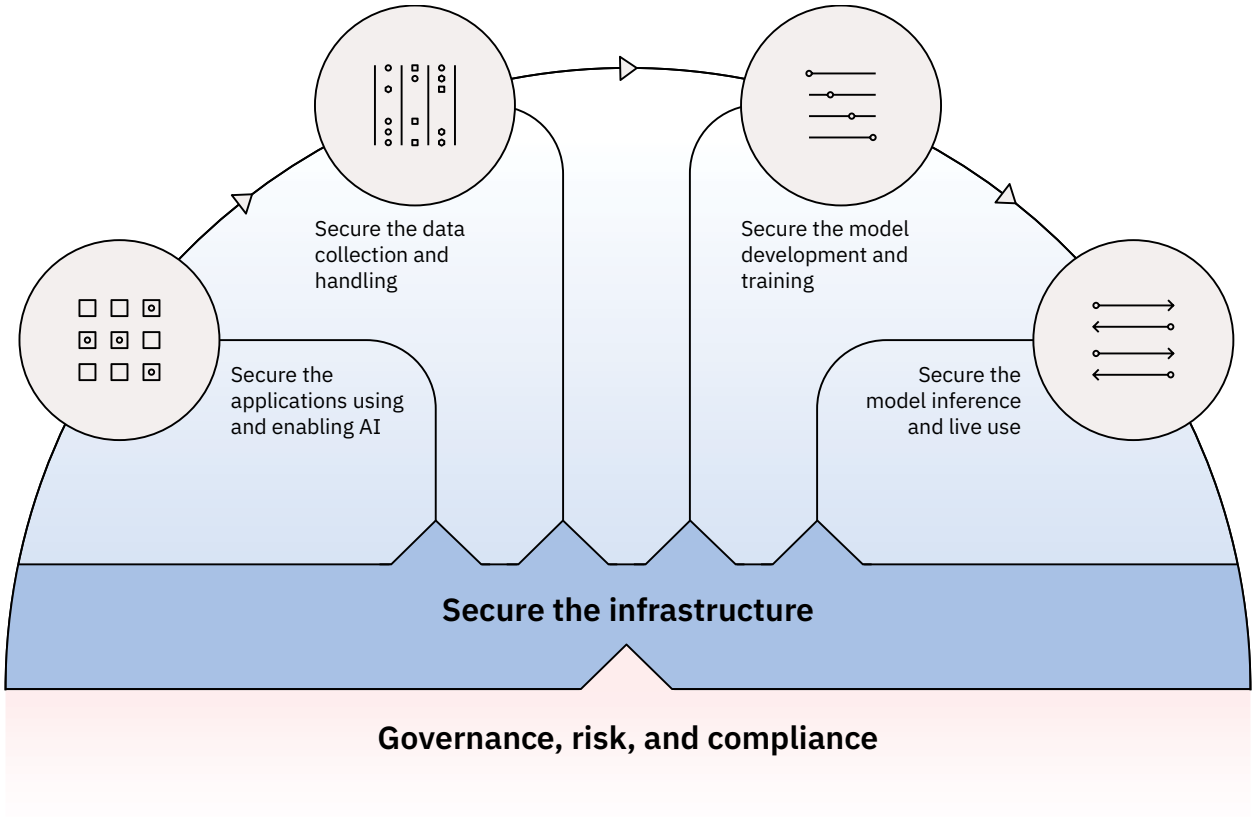


Three AI enablement models, three risk profiles

A simple framework outlines an effective approach to securing the AI pipeline—starting with updating governance, risk, and compliance (GRC) strategies (see Figure 4). Getting these principles right from the beginning—as core design considerations—can accelerate innovation. A governance and design-oriented approach to generative AI is particularly important in light of emerging AI regulatory guidance such as the EU AI Act (see Perspective, "A glimpse into new and proposed AI regulations around the world").[14] Those who integrate and embed GRC capabilities in their AI initiatives can differentiate themselves while also clearing their path to value, capitalizing on investments knowing they are building on a solid foundation.

FIGURE 4

Securing the AI value stream starts with updating risk and governance models.



Source: IBM Security.

Perspective

A glimpse into new and proposed AI regulations around the world[15]

AI regulations are evolving as quickly as gen AI models and are being established at virtually all levels of government. Organizations can look to automated AI governance tools to help manage compliance with changing policy requirements. A sampling of regulations includes:

Europe
- EU AI Act

US
- Maintaining American Leadership in AI Executive Order
- Promoting the Use of Trustworthy AI in the Federal Government Act Executive Order
- AI Training Act
- National AI Initiative Act

Canada
- AI and Data Act
- Directive on Automated Decision-Making

Brazil
- AI Bill

China
- Algorithmic Recommendations Management Provisions
- Ethical Norms for New Generation AI
- Opinions on Strengthening the Ethical Governance of Science and Technology
- Draft Provisions on Deep Synthesis Management
- Measures for the Management of Generative AI Services

Japan
- Guidelines for Implementing AI Principles
- AI Governance in Japan Ver.1.1

India
- Digital India Act

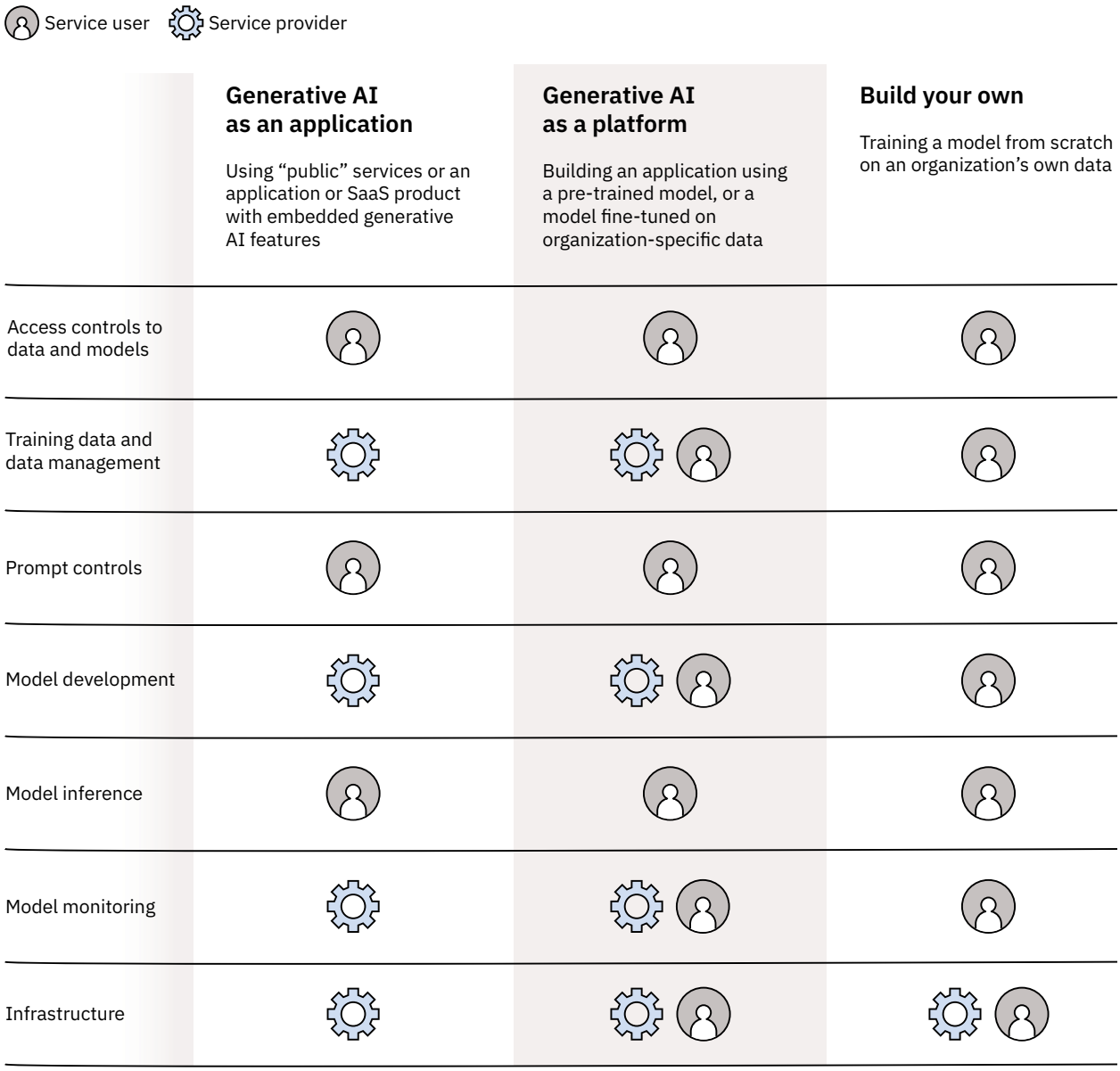
Australia
- Uses existing regulatory structures for AI oversight

Next, leaders can shift their attention to securing infrastructure and the processes comprising the AI value stream: data collection, model development, and model use. Each presents a distinct threat surface that reflects how the organization is enabling AI: using third-party applications with embedded gen AI capabilities; building gen AI solutions via a platform of pre-trained or bespoke foundation models; or building gen AI models and solutions from scratch.[16]

Each adoption route encompasses varying levels of investment, commitment, and responsibility. Working through the risks and security for each helps build resilience across the AI pipeline. While some organizations have already anchored on an adoption strategy, some are applying multiple approaches, and some may still be finding their way and formalizing their strategy. From a security perspective, what varies with each option is who is responsible for what—and how that responsibility may be shared (see Figure 5).[17]

FIGURE 5

The principles of shared responsibility extend to securing generative AI models and applications.



Source: AWS Security, IBM Security.

Using third-party applications embedded with generative AI

Organizations that are just getting started may be using consumer-focused services such as OpenAI's ChatGPT, Anthropic's Claude, or Google Gemini, or they are using an off-the-shelf SaaS product with gen AI features built in, such as Microsoft 365 or Salesforce.[18] These solutions allow organizations that have fewer investment resources to gain efficiencies from basic gen AI capabilities.

The companies providing these gen AI-enabled tools are responsible for securing the training data, the models, and the infrastructure housing the models. But users of the products are not free of security responsibility. In fact, inadvertent employee actions can induce headaches for security teams.

Similar to how shadow IT emerged with the first SaaS products and created cloud security risks, the incidence of shadow AI is growing. With employees looking to make their work lives easier with gen AI, they are complicating the organization's security posture, making security and governance more challenging.[19]

First, well-meaning staff can share private organizational data into third-party products without knowing whether the AI tools meet their security needs. This can expose sensitive or privileged data, leak proprietary data that may be incorporated into third-party models, or expose data artifacts that could be vulnerable should the vendor experience a cyber incident or data breach.[20]

Second, because the security team is unaware of the usage, they can't assess and mitigate the risks.[21] Third-party software—whether or not sanctioned by the IT/IS team—can introduce vulnerabilities because the underlying gen AI models can host malicious functionality such as trojans and backdoors.[22] One study found that 41% of employees acquired, modified, or created technology without their IT/IS team's knowledge—and predicts this percentage will climb to 75% over the next three years, exacerbating the problem.[23]

Key security considerations include:

- Have you established and communicated policies that address use of certain organizational data (confidential, proprietary, or PII) within public models and third-party applications?
- Do you understand how third parties will use data from prompts (inputs/outputs) and whether they will claim ownership of that data?
- Have you assessed the risks of third-party services and applications and know which risks they are responsible for managing?
- Do you have controls in place to secure the application interface and monitor user activity, such as the content and context of prompt inputs/outputs?

Using a platform to build generative AI solutions

Training foundation models and LLMs for generative AI applications demands tremendous infrastructure and computing resources—often beyond what most organizations can budget. Hyperscalers are stepping in with platforms that allow users to tap into a choice of pre-trained foundation models for building gen AI applications more specific to their needs. These models are trained on a large, general-purpose data set, capturing the knowledge and capabilities learned from a broad range of tasks to improve performance on a specific task or set of tasks. Pre-trained models can also be fine-tuned for a more specific task using a smaller amount of an organization's data, resulting in a new specialized model optimized around distinct use cases, such as industry-specific requirements.[24]

The open-source community is also democratizing gen AI with an extensive library of pre-trained LLMs. The most popular of these—such as Meta's Llama and Mistral AI—are also available via general-purpose gen AI platforms (see Perspective, "Risk or reward? Adopting open-source models").

Platforms offer the advantage of having some security and governance capabilities baked in. For example, infrastructure security is shared with the vendor, similar to any cloud infrastructure agreement. Perhaps the organization's data already resides with a specific cloud provider, in which case fine-tuning the model may be as simple as updating configurations and API calls. Additionally, a catalogue of enhanced security products and services is available to complement or replace the organization's own (see case study, "EVERSANA and AWS advance artificial intelligence apps for the life sciences industry").

However, when organizations build gen AI applications integrated with pre-trained or fine-tuned models, their security responsibilities grow considerably compared to using a third-party SaaS product. Now they must tackle the unique threats to foundation models and LLMs referenced in part one of this report. Risks to training data as well as the model development and inference fall squarely on their radar. Applying the principles of ModelOps and MLSecOps (machine learning security operations) can help organizations secure their gen AI applications.[25]

Key security considerations include:

- Have you conducted threat modeling to understand and manage the emerging threat vectors?
- Have you identified open-source and widely used models that have been thoroughly scanned for vulnerabilities, tested, and vetted?
- Are you managing training data workflows, such as using encryption in transit and at rest, and tracking data lineage?
- How do you protect training data from poisoning exploits that could introduce inaccuracies or bias and compromise or change the model's behavior?
- How do you harden security for API and plug-in integrations to third-party models?
- How do you monitor models for unexpected behaviors, malicious outputs, and security vulnerabilities that may appear over time?
- Are you managing access to training data and models using robust identity and access management practices, such as role-based access control, identity federation, and multifactor authentication?
- Are you managing compliance with laws and regulations for data privacy, security, and responsible AI use?

Case study

EVERSANA and AWS advance artificial intelligence apps for the life sciences industry[26]

Given regulatory requirements, life sciences companies need generative AI solutions that combine security, compliance, and scalability. EVERSANA, a leading provider of commercial services to the global life sciences industry, is turning to AWS to accelerate gen AI use cases across the life sciences industry. The objective is to harness the power of gen AI to help pharmaceutical and life science manufacturers drive efficiencies and create business value while improving patient outcomes.

EVERSANA will apply its digital and AI innovation capabilities coupled with Amazon Bedrock managed gen AI services to leverage best-of-breed foundation models. EVERSANA maintains full control over the data it uses to tailor foundation models and can customize guardrails based on its application requirements and responsible AI policies. In its first application—in partnership with AWS and TensorIoT, the team sought to automate processes associated with medical, legal, and regulatory (MLR) content approvals.

EVERSANA's strategy to leverage gen AI to solve complex challenges for life sciences companies is part of what EVERSANA calls "pharmatizing AI." Jim Lang, chief executive officer at EVERSANA, explained, "Pharmatizing AI in the life sciences industry is about leveraging technology to optimize and accelerate common processes that are desperate for innovation and transformation." This approach has led to streamlining critical processes from months to weeks. EVERSANA anticipates that once it automates its MLR capabilities, it can further improve time-to-approval from weeks to mere days.

EVERSANA anticipates automation of MLR processes can improve approval time from weeks to days.

Perspective

Risk or reward? Adopting open-source models[27]

In contrast to proprietary LLMs that can only be used by customers who purchase a license from the provider, open-source LLMs are free and available for anyone to access. They can be used, modified, and distributed with far greater flexibility than proprietary models.

Designed to offer transparency and interoperability, open-source LLMs allow organizations with minimal machine learning skills to adapt gen AI models for their own needs—and on their own cloud or on-premises infrastructure. They also help offset concerns about the risk of becoming overly reliant on a small number of proprietary LLMs.

Risks with using open-source models are similar to proprietary models, including hallucinations, bias, and accountability issues with the training data. But the trait that makes open source popular—the community approach to development—can also be its greatest vulnerability as hackers can more easily manipulate core functionality for malicious purposes. These risks can be mitigated by adopting security hygiene practices as well as software supply chain and data governance controls.

Open-source LLMs allow organizations with minimal machine learning skills to adapt gen AI models for their own needs.

Building your own generative AI solutions

A few large organizations with deep pockets are building and training LLMs—and smaller, more tailored language models (SLMs)[28]—from scratch based solely on their data. Hyperscaler tools are helping accelerate the training process, while the organization owns every aspect of the model. This can afford them performance advantages as well as more precise results.[29]

In this scenario, on top of the governance and risk management outlined for applications based on pre-trained and fine-tuned models, the organization's own data security posture takes on greater importance. As the organization's data is now incorporated into the AI model itself, responsible AI becomes essential to reducing risk exposure.

Being the primary source for AI training data, organizations are responsible for making sure that data—and the outcomes based on it—can be trusted. That means protecting the source data following strict data security practices (see Perspective, "Why responsible AI starts with security ABCs"). And it means protecting the models from being compromised or exploited by malicious actors. Access controls, encryption, and threat detection systems are critical pieces in preventing data from being manipulated. The trustworthiness of an AI solution may be measured by its ability to offer unbiased, accurate, and ethical responses.

If organizations do not practice responsible AI, they risk damage to their brands from faulty—even dangerous—output from their gen AI models. Despite these risks, fewer than 20% of executives say they are concerned about a potential liability for erroneous outputs from gen AI. In other IBM IBV research, only 30% of respondents said they are validating the integrity of gen AI outputs.[30]

If secure and trustworthy data is the basis for value generation—and much of our research indicates it is—leaders should focus on the security implications of (ir)responsible AI.[31] Doing so can highlight the various ways AI models may be manipulated. In the absence of bias or explainability controls, such manipulation can be hard to recognize. This is why organizations need a strong foundation in governance, risk, and compliance.

As an extension of the organization's data security posture, software supply chain security also becomes more consequential when creating LLMs. These models are built on top of complex software stacks that include multiple layers of software dependencies, libraries, and frameworks. Each of these components can introduce vulnerabilities that can be exploited by attackers to compromise the integrity of the AI model or the underlying data.

Unfortunately, adoption of software supply chain security best practices is still nascent at many organizations, according to recent IBM IBV research. For example, only 29% of executives indicated they have adopted DevSecOps principles and practices to secure their software supply chain, and only 32% have implemented continuous monitoring capabilities for their software suppliers.[32] Both practices are vital to helping prevent cyber incidents throughout the software supply chain.

Key security considerations include:

- Do you need to bolster data security practices to help prevent theft and manipulation and support responsible AI?
- How can you shore up third-party software security awareness and practices; for example, ensuring that zero-trust principles are in place?
- Do you require procurement teams to check supplier contracts for security vulnerability controls and risk-related performance measures?

Perspective

Why responsible AI starts with security ABCs

As AI moves from experimentation into production, the ABCs of security—awareness, behavior, and culture—become even more important for helping ensure responsible AI. For AI to be designed, developed, and deployed with good intent for the benefit of society, trust is an imperative.[33]

Consistent with many emerging technologies, well-informed employees and partners can be an asset—especially in light of new multimodal and rich-media-based phishing tactics enabled by gen AI. Enhancing employee awareness of the new risks leads to proactive behaviors and, over time, a more robust security culture.

As AI solutions become more integral to operations, a standard practice should be to communicate new functionality and associated security controls to employees, while reiterating the policies in place to protect proprietary and personal data. Established controls should be updated to address new threats, with the core principles of zero trust and least privilege limiting lateral movement. Emphasizing a sense of ownership about security outcomes can reinforce security as a common, shared endeavor connecting virtually all stakeholders and partners. Responsible AI is about more than policies—it's a commitment to safeguard the trust that's critical to the organization's continuing success.

Enhancing employee awareness of new risks from AI can lead to proactive behaviors and, over time, a more robust security culture.

Part three

The leadership dilemma—generative AI requires what organizations have least

Developing and securing generative AI solutions requires capacity, resources, and skills—the very things organizations don't have enough of.[34] In fragmented IT environments, security takes on higher levels of complexity that require even more capacity, resources, and skills. Leaders quickly find themselves in a dilemma.

AI-enhanced tools

AI-powered security products can bridge the skills gap by freeing overworked staff from time-consuming tasks. This allows them to focus on more complex security issues that require expertise and judgment. By optimizing time and resources, AI effectively adds capacity and skills. With improved insights, productivity, and economies of scale, organizations can adopt a more preventive and proactive security posture. Indeed, leading security AI adopters cut the time to detect incidents by one-third and the costs of data breaches by at least 18%.[35] New capabilities are also emerging that automate management of compliance within a rapidly changing regulatory environment.

The shift to AI security tools is consistent with how cybersecurity demand is changing. While the market for AI security products is expected to grow at a CAGR of nearly 22% over the next five years, providers are focusing on developing consolidated security software solutions. To facilitate better efficiency and governance, solution providers are rationalizing their toolsets and streamlining data analysis.[36] This more holistic approach to security enhances visibility across the operations lifecycle—something 53% of executives are expecting to gain from gen AI.

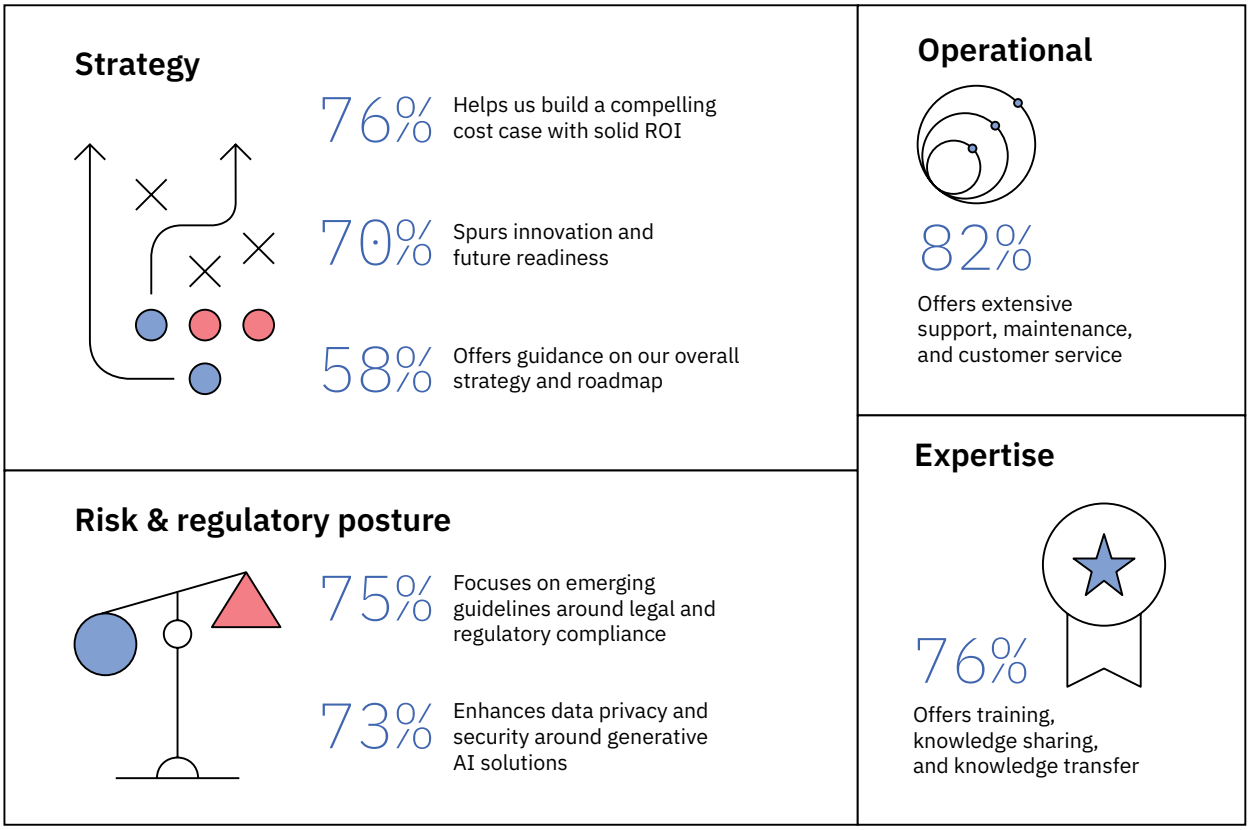
AI-experienced partners

Business partners can also help close security skills gaps. Just as with the transition to cloud, partners can assist with assessing needs and managing security outcomes. Amid the ongoing security talent shortage that's exacerbated by a lack of AI skills, organizations are seeking partners that can facilitate training, knowledge sharing, and knowledge transfer (76%). They are also looking for gen AI partners to provide extensive support, maintenance, and customer service (82%). Finally, they are choosing partners that can guide them across the evolving legal and regulatory compliance landscape (75%).

Executives are also in search of partners to help with strategy and investment decisions (see Figure 6). With around half (47%) saying they are uncertain about where and how much to invest, it's no surprise that three-quarters (76%) want a partner to help build a compelling cost case with solid ROI. More than half also seek guidance on an overall strategy and roadmap.

FIGURE 6

Executives are turning to partners to help deliver and support generative AI security solutions.

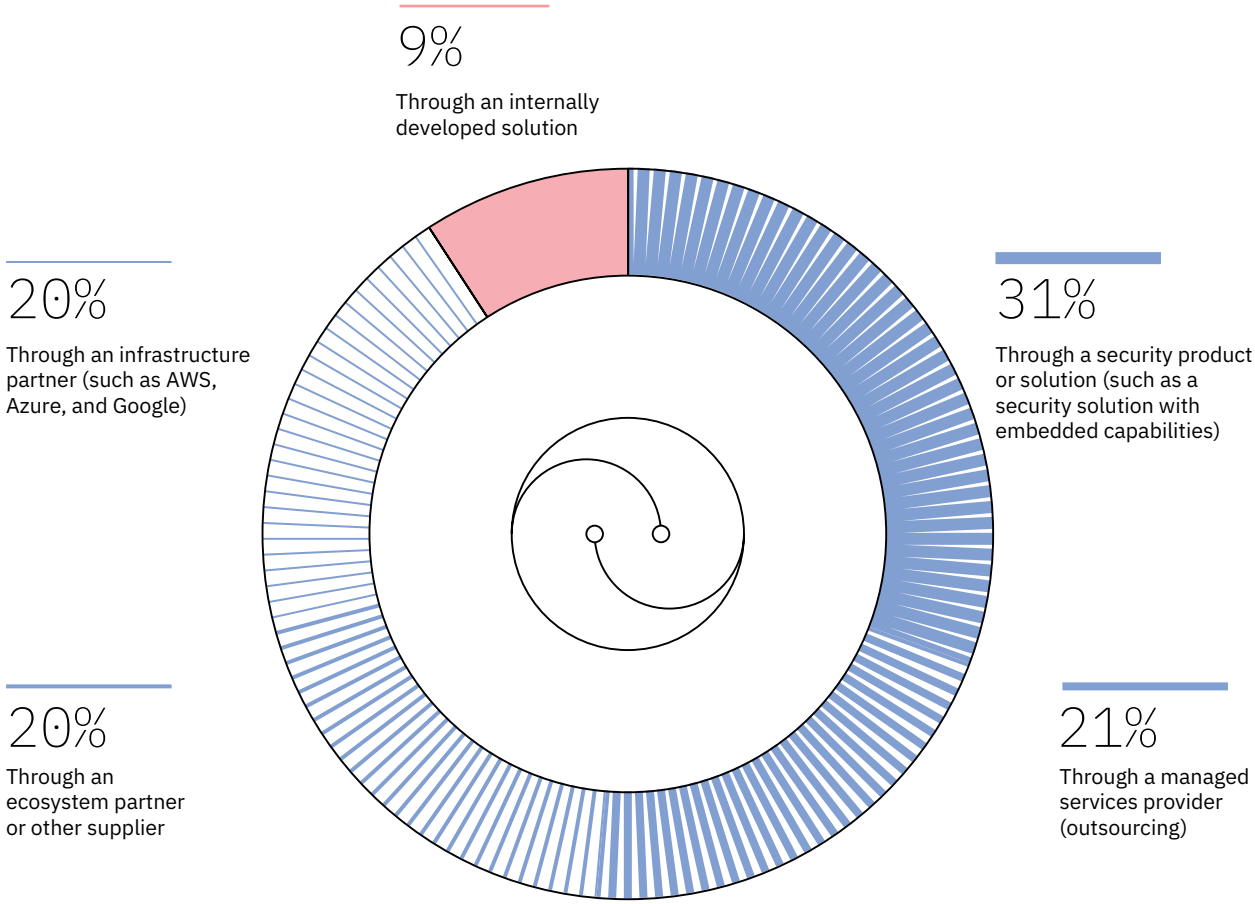


Q. How important are these when choosing a partner for your generative AI security needs?

Our results indicate that most organizations are turning to partners to enable generative AI for security. While many respondents are purchasing security products or solutions with gen AI capabilities, nearly two-thirds of their security generative AI capabilities are coming through some type of partner—managed services, ecosystem/supplier, or hyperscaler (see Figure 7). Similar to cloud adoption, leaders are looking to partners for comprehensive security support—whether that's informing and advising about generative AI or augmenting their delivery and support capabilities.

FIGURE 7

More than 90% of security gen AI capabilities are coming from third-party products or partners.



Q. How are you enabling generative AI for security capabilities?
Note: percentages do not add to 100% due to rounding.

Action guide

Whether just starting to experiment with generative AI, building models on your own, or somewhere in between, the following guidance can help organizations secure their AI pipeline. These recommendations are intended to be cross-functional, facilitating engagement across security, technology, and business domains.

01

Assess

- Define an AI security strategy that aligns with the organization's overall AI strategy.
- Ask how your organization is using AI today—for which use cases, in what applications, through which service providers, and serving which user cohorts. Once you answer these questions, then quantify the associated sources of risk.
- Evaluate the maturity of your core security capabilities, including infrastructure security, data security, identity and access management practices, threat detection and incident response, regulatory compliance, and software supply chain management. Identify where you must be better to support the demands of AI.
- Decide where partners can supplement and complement your security capabilities and define how responsibilities will be shared.
- Uncover security gaps in AI environments using risk assessment and threat modeling. Determine how policies and controls need to be updated to address emergent threat vectors driven by generative AI.

02

Implement

- Establish AI governance working with business units, risk, data, and security teams.
- Prioritize a secure-by-design approach across the ML and data pipeline to drive safe software development and implementation.
- Manage risk, controls, and trustworthiness of AI model providers and data sources.
- Secure AI training data in line with current data privacy and regulatory guidelines, and adopt new guidelines when published.
- Secure workforce, machine, and customer access to AI apps and subsystems from anywhere.

03

Monitor

- Evaluate model vulnerabilities, prompt injection risks, and resiliency with adversarial testing.
- Perform regular security audits, penetration testing, and red-teaming exercises to identify and address potential vulnerabilities in the AI environment and connected apps.

04

Educate

- Review cyber hygiene practices and security ABCs (awareness, behaviors, and culture) across your organization.
- Conduct persona-based cybersecurity awareness activities and education, particularly as they relate to AI as a new threat surface. Target all stakeholders involved in the development, deployment, and use of AI models, including employees using AI-powered tools.

Authors

Clarke Rodgers

Director, AWS Enterprise Strategy
linkedin.com/in/clarkerodgers/
rodgclar@amazon.com

Moumita Saha

Senior Security Partner Solutions Architect, AWS
linkedin.com/in/moumita-saha/
moumis@amazon.com

Dimple Ahluwalia

Vice President and Global Managing Partner
IBM Consulting Cybersecurity Services
linkedin.com/in/dimple-ahluwalia-08b70/
Dimple.Ahluwalia@ibm.com

Kevin Skapinetz

Vice President, Strategy and Product Management
IBM Security
linkedin.com/in/kskap/
kskap@us.ibm.com

Gerald Parham

Global Research Leader, Security and CIO
IBM Institute for Business Value
linkedin.com/in/gerryparham/
gparham@us.ibm.com

Expert contributors

Heather Deguzman

Senior Executive Marketing Manager, Content
Amazon Web Services
linkedin.com/in/heather-deguzman/
deguheat@amazon.com

Ryan Dougherty

Program Director, Product Management,
Emerging Security Technology
IBM Security
linkedin.com/in/ryan-dougherty-2a781b1/
rdougherty@us.ibm.com

Kevin Gray

Brand and Content Strategy, Security
IBM Marketing
linkedin.com/in/kevin-gray-aba70870/
Kevin.Gray2@ibm.com

Sam Hector

Product Manager, Emerging Security Technology
IBM Security
linkedin.com/in/samhector/
samhector@uk.ibm.com

Michael Massimi

Global Principal, Cloud Security Services for
AWS Consulting
IBM Consulting
linkedin.com/in/michael-massimi-b4b3b21/
mmassimi@us.ibm.com

George Mina

Program Director, Product Management
Emerging Security Technology and Ventures
IBM Security
linkedin.com/in/gmina1/
geemin11@us.ibm.com

Dinesh Nagarajan

Global Cyber Trust Partner, Portfolio Leader
IBM Consulting Cybersecurity Services
linkedin.com/in/dineshnagarajan/
Dinesh.Nagarajan@uk.ibm.com

Georgia Prassinos

Security Communications
IBM Marketing
linkedin.com/in/georgiaprassinos/
gprassinos@ibm.com

Jeremy Testerman

Manager, AI and Platform Product Management
IBM Consulting
linkedin.com/in/jeremy-testerman-4a560a2/
jtester@us.ibm.com

Srini Tummalapenta

Distinguished Engineer and CTO
Master Inventor
IBM Consulting Cybersecurity Services
linkedin.com/in/srinivastummalapenta/
stummala@us.ibm.com

IBM Institute for Business Value editorial and design team

Sara Aboulsohn, Visual Designer
Kris Biron, Visual Designer
Joanna Wilkins, Editorial Lead

Study methodology and approach

In Q3 2023, the IBM Institute for Business Value partnered with Oxford Economics to survey 200 executives about their generative AI security strategy and enablement. Respondents are based in the US and responsible for operations at either US-based organizations or multinational organizations with a significant US presence. Respondents include CEOs, CISOs, CIOs, and Chief Data Officers.

Respondents were screened for several inclusion criteria. They indicated whether they are either moderately familiar or very familiar with generative AI. Respondent organizations are either in the piloting or implementation phases of generative AI. Respondents described their familiarity with their organization's security spending and investments as either "aware and consistently involved" or "working on projects and influencing investments."

Respondents represent the following industries: consumer banking, consumer products, energy and utilities, financial markets, government (federal), government (state/provincial), healthcare providers, industrial manufacturing (industrial products), insurance, IT services, life sciences/pharmaceuticals, manufacturing (non-industrial), oil and gas, retail, telecommunications, transportation, and travel.

About the AWS-IBM Security partnership

IBM is an AWS Premier Tier Consulting Partner, including three security competencies and a total of 16 AWS competencies across IBM Technology and IBM Consulting. Together, IBM and AWS bring fast, security-rich, open software capabilities to the cloud platform for more than one million customers every day. The power of cloud-native AWS capabilities, combined with 50+ IBM solutions available on AWS Marketplace, enables clients to access AI-powered IBM Software with turnkey delivery and integration. For more information, visit https://www.ibm.com/aws/security

Related reports

The CEO's guide to generative AI: Cybersecurity

The CEO's guide to generative AI: Cybersecurity.
IBM Institute for Business Value. October 2023.
https://ibm.co/ceo-generative-ai-cybersecurity

Data security as business accelerator?

Data security as business accelerator?
The unsung hero driving competitive advantage.
IBM Institute for Business Value and Amazon Web Services. June 2023. https://ibm.co/data-security

AI and automation for cybersecurity

AI and automation for cybersecurity: How leaders succeed by uniting technology and talent.
IBM Institute for Business Value. June 2022.
https://ibm.co/ai-cybersecurity

About Research Insights

Research Insights are fact-based strategic insights for business executives on critical public- and private-sector issues. They are based on findings from analysis of our own primary research studies. For more information, contact the IBM Institute for Business Value at iibv@us.ibm.com.

IBM Institute for Business Value

For two decades, the IBM Institute for Business Value has served as the thought leadership think tank for IBM. What inspires us is producing research-backed, technology-informed strategic insights that help leaders make smarter business decisions.

From our unique position at the intersection of business, technology, and society, we survey, interview, and engage with thousands of executives, consumers, and experts each year, synthesizing their perspectives into credible, inspiring, and actionable insights.

To stay connected and informed, sign up to receive IBV's email newsletter at ibm.com/ibv. You can also find us on LinkedIn at https://ibm.co/ibv-linkedin.

The right partner for a changing world

At IBM, we collaborate with our clients, bringing together business insight, advanced research, and technology to give them a distinct advantage in today's rapidly changing environment.

Notes and sources

1 Britton, Mike. "Uncovering AI-Generated Email Attacks: Real-World Examples from 2023." *Abnormal Blog.* December 19, 2023. https://abnormalsecurity.com/blog/2023-ai-generated-email-attacks; Chen, Heather and Kathleen Magramo. "Finance worker pays out $25 million after video call with deepfake 'chief financial officer.'" CNN. February 4, 2024. https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html; Kharpal, Arjun. "Samsung bans use of A.I. like ChatGPT for employees after misuse of the chatbot." CNBC. May 2, 2023. https://www.cnbc.com/2023/05/02/samsung-bans-use-of-ai-like-chatgpt-for-staff-after-misuse-of-chatbot.html

2 *X-Force Threat Intelligence Index 2024.* IBM Security. February 2024. https://www.ibm.com/reports/threat-intelligence

3 Sabin, Sam. "Generative AI puts GPU security in the spotlight." *Axios.* March 22, 2024. https://www.axios.com/2024/03/22/generative-ai-chips-gpu-security; Repiso, Jorge. "The players set to shape the AI landscape in 2024." Digitalis. March 2024. https://digitalis.com/news/the-players-set-to-shape-the-ai-landscape-in-2024/

4 IBM Institute for Business Value survey of 2,500 global, cross-industry executives on AI adoption. 2024. Unpublished data.

5 Isola, Laurie. "How cybercriminals are using gen AI to scale their scams." *Okta Blog.* January 4, 2024. https://www.okta.com/blog/2024/01/how-cybercriminals-are-using-gen-ai-to-scale-their-scams/; Allan, Katy. "Cybercriminals are creating a darker side to AI." *Cyber Magazine.* October 24, 2024. https://cybermagazine.com/articles/cybercriminals-are-creating-a-darker-side-to-ai; Kelley, Daniel. "Exploring the World of AI Jailbreaks." *SlashNext Blog.* September 12, 2023. https://slashnext.com/blog/exploring-the-world-of-ai-jailbreaks/; "6 Prompts You Don't Want Employees Putting in Microsoft Copilot." *BleepingComputer.* April 3, 2024. https://www.bleepingcomputer.com/news/security/6-prompts-you-dont-want-employees-putting-in-microsoft-copilot/; Goodin, Dan. "Thousands of servers hacked in ongoing attack targeting Ray AI framework." *Ars Technica.* March 27, 2024. https://arstechnica.com/security/2024/03/thousands-of-servers-hacked-in-ongoing-attack-targeting-ray-ai-framework/

6 "Series 2 of 10: Cybersecurity Architecture Fundamentals." IBM Technology YouTube video. July 2023. https://www.youtube.com/watch?v=EqNe55IzjAw

7 "What's an AI pipeline?" *Squark Blog.* Accessed April 11, 2024. https://squarkai.com/whats-an-ai-pipeline/

8 "Email Attack Trends: How phishing attacks are becoming more sophisticated and harder to identify." *Darktrace Blog.* March 20, 2024. https://darktrace.com/blog/email-attack-trends-how-phishing-attacks-are-becoming-more-sophisticated-and-harder-to-identify

9 *X-Force Threat Intelligence Index 2024.* IBM Security. February 2024. https://www.ibm.com/reports/threat-intelligence

10 "Email Attack Trends: How phishing attacks are becoming more sophisticated and harder to identify." *Darktrace Blog.* March 20, 2024. https://darktrace.com/blog/email-attack-trends-how-phishing-attacks-are-becoming-more-sophisticated-and-harder-to-identify

11 McCurdy, Chris, Sholmi Kramer, Gerald Parham, and Jacob Dencik, PhD. *Prosper in the cyber economy: Rethinking cyber risk for business transformation.* November 2022. Unpublished data.

12 Hector, Sam. "Mapping attacks on generative AI to business impact." *Security Intelligence.* January 30, 2024. https://securityintelligence.com/posts/mapping-attacks-generative-ai-business-impact/

13 "The Value (and Threat) of Generative AI for Security Teams." AWS podcast. Accessed April 22, 2024. https://aws.amazon.com/podcasts/cwl-the-value-and-threat-of-generative-ai-for-security-teams/

14 The EU Artificial Intelligence Act website. Accessed April 11, 2024. https://artificialintelligenceact.eu/

15 Ponomarov, Kostiantyn. "Global AI Regulations Tracker: Europe, Americas & Asia-Pacific Overview." Legal Nodes. March 20, 2024. https://legalnodes.com/article/global-ai-regulations-tracker

16 Saner, Matt and Mike Lapidakis. "Securing generative AI: An introduction to the Generative AI Security Scoping Matrix." *AWS Security Blog.* October 19, 2023. https://aws.amazon.com/blogs/security/securing-generative-ai-an-introduction-to-the-generative-ai-security-scoping-matrix/

17 Manral, Vishwas. "Generative AI: Proposed Shared Responsibility Model." *Cloud Security Alliance Blog.* July 28, 2023. https://cloudsecurityalliance.org/blog/2023/07/28/generative-ai-proposed-shared-responsibility-model

18 Ibid.

19 Salvin, Steve. "What managers should know about the secret threat of employees using 'shadow AI.'" *Fast Company.* October 26, 2023. https://www.fastcompany.com/90972657/what-managers-should-know-about-the-secrets-threat-of-employees-using-shadow-ai

20 Ibid.

21 Ibid.

22 Hector, Sam. "Mapping attacks on generative AI to business impact." *Security Intelligence.* January 30, 2024. https://securityintelligence.com/posts/mapping-attacks-generative-ai-business-impact/

23 "Gartner Unveils Top Eight Cybersecurity Predictions for 2023-2024." Gartner Newsroom. March 28, 2023. https://www.gartner.com/en/newsroom/press-releases/2023-03-28-gartner-unveils-top-8-cybersecurity-predictions-for-2023-2024

24 Saner, Matt and Mike Lapidakis. "Securing generative AI: An introduction to the Generative AI Security Scoping Matrix." *AWS Security Blog.* October 19, 2023. https://aws.amazon.com/blogs/security/securing-generative-ai-an-introduction-to-the-generative-ai-security-scoping-matrix/

25 Kerner, Sean Michael. "Exclusive: What will it take to secure gen AI? IBM has a few ideas." *VentureBeat.* January 25, 2024. https://venturebeat.com/ai/exclusive-what-will-it-take-to-secure-gen-ai-ibm-has-a-few-ideas/; Pariseau, Beth. "Meet MLSecOps: Industry calls for new measures to secure AI." *TechTarget News.* September 13, 2023. https://www.techtarget.com/searchitoperations/news/366552019/Meet-MLSecOps-industry-calls-for-new-measures-to-secure-AI

26 "EVERSANA & Amazon Web Services to 'Pharmatize' Artificial Intelligence across the Life Sciences Industry." EVERSANA news release. July 24, 2023. https://www.eversana.com/2023/07/24/eversana-amazon-web-services-to-pharmatize-artificial-intelligence-across-the-life-sciences-industry/; Amazon Bedrock website. Accessed April 11, 2024. https://aws.amazon.com/bedrock/; "EVERSANA Collaborates with AWS and TensorIoT to Automate the Regulatory Review Process." AWS case study. Accessed April 12, 2024. https://aws.amazon.com/partners/success/eversana-tensoriot/

27 "Open source large language models: Benefits, risks and types." *IBM Think Blog.* September 27, 2023. https://www.ibm.com/blog/open-source-large-language-models-benefits-risks-and-types/

28 Javaheripi, Mojan and Sébastien Bubeck. "Phi-2: The surprising power of small language models." *Microsoft Research Blog.* December 12, 2023. https://www.microsoft.com/en-us/research/blog/phi-2-the-surprising-power-of-small-language-models/

29 AWS Trainium website. Accessed April 11, 2024. https://aws.amazon.com/machine-learning/trainium/

30 IBM Institute for Business Value survey of 2,000 global executives responsible for supplier management, supplier sourcing, and ecosystem partner relationships. 2023. Unpublished data.

31 *IBV C-suite Series. Turning data into value: How top Chief Data Officers deliver outsize results while spending less.* IBM Institute for Business Value. March 2023. https://ibm.co/c-suite-study-cdo

32 IBM Institute for Business Value survey of 2,000 global executives responsible for supplier management, supplier sourcing, and ecosystem partner relationships. 2023. Unpublished data.

33 "What is responsible AI?" IBM website. Accessed April 11, 2024. https://www.ibm.com/topics/responsible-ai

34 "Data Suggests Growth in Enterprise Adoption of AI is Due to Widespread Deployment by Early Adopters, But Barriers Keep 40% in the Exploration and Experimentation Phases." IBM Newsroom. January 10, 2024. https://newsroom.ibm.com/2024-01-10-Data-Suggests-Growth-in-Enterprise-Adoption-of-AI-is-Due-to-Widespread-Deployment-by-Early-Adopters; Huang, Hugo. "What CEOs Need to Know About the Costs of Adopting GenAI." *Harvard Business Review.* November 15, 2023. https://hbr.org/2023/11/what-ceos-need-to-know-about-the-costs-of-adopting-genai

35 Fisher, Lisa and Gerald Parham. *AI and automation for cybersecurity: How leaders succeed by uniting technology and talent.* IBM Institute for Business Value. May 2022. https://ibm.co/ai-cybersecurity

36 "Artificial Intelligence in Cybersecurity Market by Offering (Hardware, Solution, and Service), Security Type, Technology (ML, NLP, Context-Aware and Computer Vision), Application (IAM, DLP, and UTM), Vertical and Region – Global Forecast to 2028." Markets and Markets. December 2023. https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-ai-cyber-security-market-220634996.html; "What is a Cybersecurity Platform?" Trend Micro. Accessed April 11, 2024. https://www.trendmicro.com/en_us/what-is/cybersecurity-platform.html

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America | May 2024

2L73BYB4-USEN-00