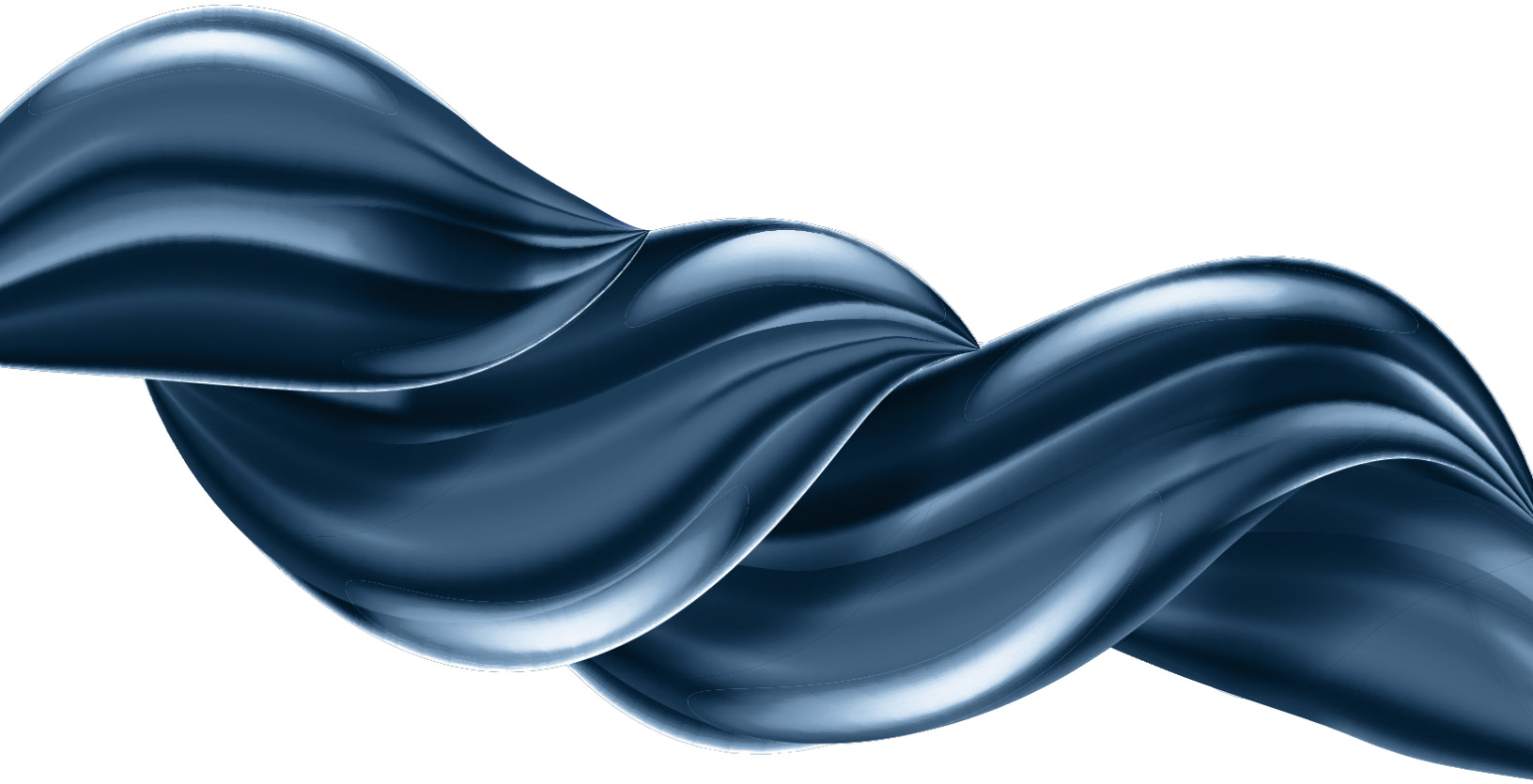
McKinsey & Company

McKinsey on Risk & Resilience

Managing risks to achieve resilience



Number 16, March 2024

The articles in *McKinsey on Risk & Resilience* are written by risk experts and practitioners from McKinsey's Risk & Resilience Practice and other firm practices. This publication offers readers insights into value-creating strategies and the translation of those strategies into company performance.

This issue, and future issues, are available to registered users online at McKinsey.com. Comments and requests for copies or for permissions to republish an article can be sent via email to McKinsey_Risk@McKinsey.com.

Cover image:
© VectorUp/Getty Images

Editorial Board:
Bob Bartels, Oliver Bevan, Joseba Eceiza, Justin Greis, Carina Kofler, Andreas Kremer, Mihir Mysore, Thomas Poppensieker, Sebastian Schneider, Lorenzo Serino, Marco Vettori, David Weidner

External Relations, Global Risk & Resilience Practice:
Bob Bartels

Editor: David Weidner

Contributing editors:
Charlie Barthold, Larry Kanter, Joanna Pachner

Art Direction and Design:
LEFF

Data Visualization:
Richard Johnson, Matt Perry, Jonathon Rivait, Jessica Wang

Managing Editor:
Heather Byer

Editorial Production:
Mark Cajigao, Nancy Cohn, Roger Draper, Ramya D'Rozario, Mary Gayen, Gwyn Herbein, Drew Holzfeind, LaShon Malone, Pamela Norton, Katrina Parker, Kanika Punwani, Charmaine Rice, Dana Sand, Katie Shearer, Regina Small, Maegan Smith, Sarah Thuerk, Sneha Vats, Pooja Yadav

McKinsey Global Publications

Publisher: Raju Narisetti

Global Editorial Director and Deputy Publisher:
Lucia Rahilly

Global Publishing Board of Editors: Roberta Fusaro, Bill Javetski, Lucia Rahilly, Mark Staples, Rick Tetzeli, Monica Toriello

Copyright © 2024 McKinsey & Company. All rights reserved.

This publication is not intended to be used as the basis for trading in the shares of any company or for undertaking any other complex or significant financial transaction without consulting appropriate professional advisers.

No part of this publication may be copied or redistributed in any form without the prior written consent of McKinsey & Company.

Contents

Introduction

Welcome to the latest issue of *McKinsey on Risk & Resilience*, now a quarterly publication featuring the latest trends, research, insights, and best practices related to financial and nonfinancial risks, and how to build and strengthen business resilience in an ever-evolving world.

Our move to a more frequent quarterly publishing cadence is aimed at providing risk leaders with the latest data and insights in a timelier and more relevant manner.

In the past few years, organizations have faced various risks that have had significant effects on their operations and performance. These risks encompass a wide range of areas and can vary depending on the industry and specific circumstances of each organization.

Emerging technology. From modeling analytics to automating manual tasks and synthesizing unstructured content, generative AI (gen AI) is already changing how organizations operate. And this is just the start: McKinsey has estimated the annual potential of gen AI to the banking sector alone at $200 billion to $340 billion in new value. Use cases and successful best practices already exist, and organizations, along with their risk and compliance functions, are adapting to these challenges and opportunities.

Geopolitical tensions. Geopolitical tensions such as trade wars and political instability pose significant risks to organizations. These tensions can disrupt global markets, increase regulatory uncertainties, and affect international trade. Organizations operating in multiple countries must navigate through these risks and adapt their strategies accordingly.

Supply chain risk. Organizations increasingly rely on complex global supply chains. Disruptions in the supply chain such as natural disasters, political conflicts, or pandemics can have severe financial consequences. The COVID-19 pandemic, for example, highlighted the vulnerability of organizations to supply chain disruptions, leading to production delays, inventory shortages, and increased costs.

Inflation and rising interest rates. Inflation and rising interest rates can significantly affect organizations' financial health. Inflation erodes purchasing power and increases costs, and rising interest rates can increase borrowing costs and affect investment decisions. Organizations need to carefully manage these risks to ensure their financial stability and profitability.

Cybersecurity threats. Given the increasing reliance on technology and digital infrastructure, organizations face growing cybersecurity risks. Cyberattacks can result in financial losses, damage to reputations, and regulatory penalties. To mitigate these risks, organizations should invest in robust cybersecurity measures and develop effective incident response plans.

While this is certainly not an exhaustive list, these risks are significant, with potential widespread impact on organizations across various industries. Organizations should be proactive in identifying and managing these risks to ensure their financial resilience and long-term success.

In this issue, we delve into the emergence of gen AI and its impact on both regulators and risk and compliance functions, as well as how to implement gen AI with speed and safety. We also look at how CEOs can mitigate compounding risks. Last, our ongoing efforts with the World Economic Forum's Resilience Consortium reveal examples of organizations in the private and public sectors putting resilience measures into action.

We hope you enjoy these articles and find in them ideas worthy of application. Let us know what you think at McKinsey_Risk@McKinsey.com and on the McKinsey Insights app.

Thomas Poppensieker
Senior partner and chair,
Global Risk & Resilience Editorial Board

Copyright © 2024 McKinsey & Company. All rights reserved.

Building a resilient tomorrow: Concrete actions for global leaders

Leaders need to move toward putting resilience into action.

© Getty Images

Since its launch in 2022, the Resilience Consortium has aimed to harmonize and reinforce resilience-building efforts across the public and private sectors. To do so, it released two comprehensive reports focusing on identifying themes and the enablers required to start building resilience.

Since opening the floor to such conversations, the consortium has shone a light on numerous remarkable examples of resilience from the private and public sectors. However, it would be unrealistic to assume that there is not much work to be done. The time to act is now, and organizations need to move from "talking the talk" to "walking the walk."

To put resilience into action, the consortium believes showcasing examples from organizations that have already embarked on this journey is vital, because it can serve as a source of inspiration for those looking to embark on or progress further along their path toward resilience. This paper provides an in-depth analysis of nine case studies across three resilience themes: climate, energy, and food; supply chain; and organizational readiness. These case studies cover the public and private sectors, impacting four continents.

The case studies represent a spectrum of initiatives that showcase the diverse approaches organizations are taking to enhance resilience and adapt to global challenges. The insights derived from them, along with the dialogues the consortium has engaged in over the last two years, have enabled the identification of seven priority actions across three pillars: building the resilience muscle with new leadership and organizational capabilities; understanding, measuring, and monitoring your organization along its entire resilience journey; and developing public–private partnerships to address challenges no one party can tackle alone. These actions are intended to serve as guiding principles for senior leaders as they strive to strengthen their ability to thrive in a risk-prone world.

Exhibit

Seven actions for leaders to build resilience within their organization

Pillar 1

Build the resilience muscle with new resilience leadership and organizational capabilities

1. Develop a new resilience leadership mindset
2. Create a resilience agenda addressing short- and longer-term risks and opportunities

Pillar 2

Understand, measure, and monitor your organization along its entire resilience journey

3. Assess your organization against a resilience framework
4. Develop methodologies to factor resilience in decision making
5. Continuously measure and communicate the resilience status to internal and external stakeholders

Pillar 3

Develop public–private partnerships to address challenges no one party can tackle alone

6. Develop new financing and insurance mechanisms to derisk resilience
7. Set up a public–private partnership machinery to promote collaboration through multiple interventions

McKinsey & Company

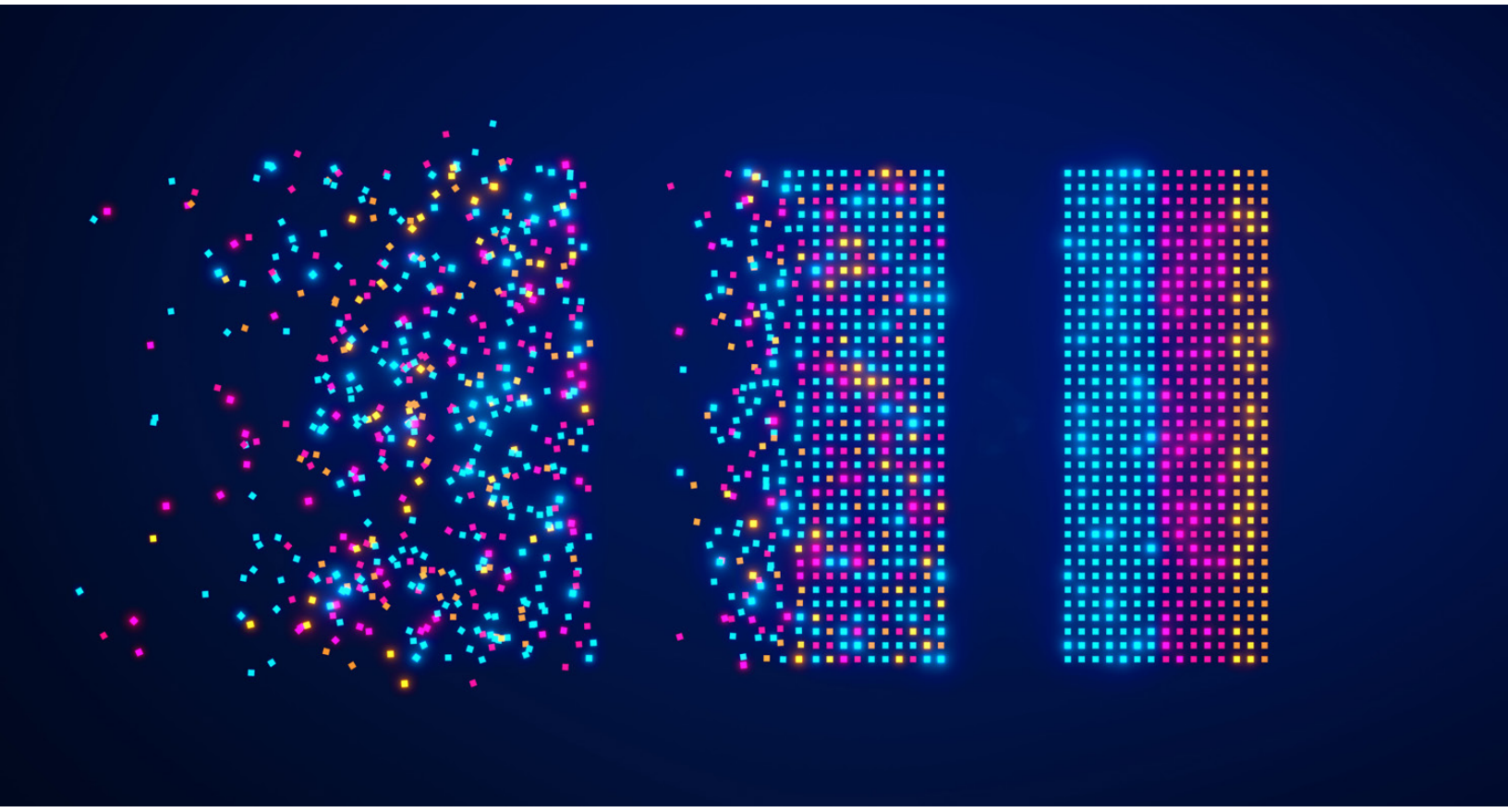
Download the full report on which this article is based, *Building a resilient tomorrow: Concrete actions for global leaders.*

Copyright © 2024 McKinsey & Company. All rights reserved.

How generative AI can help banks manage risk and compliance

In the next five years, generative AI could fundamentally change financial institutions' risk management by automating, accelerating, and enhancing everything from compliance to climate risk control.

by Rahul Agarwal, Andreas Kremer, Ida Kristensen, and Angela Luget



© Getty Images

Generative AI (gen AI) is poised to become a catalyst for the next wave of productivity gains across industries, with financial services very much among them. From modeling analytics to automating manual tasks to synthesizing unstructured content, the technology is already changing how banking functions operate, including how financial institutions manage risks and stay compliant with regulations.

It's imperative for risk and compliance functions to put guardrails around gen AI's use in an organization. However, the tech can help the functions themselves improve efficiency and effectiveness. In this article, we discuss how banks can build a flexible, powerful approach to using gen AI in risk and compliance management and identify some crucial topics that function leaders should consider.

Seizing the promise of gen AI

Gen AI has the potential to revolutionize the way that banks manage risks over the next three to five years. It could allow functions to move away from task-oriented activities toward partnering with business lines on strategic risk prevention and having controls at the outset in new customer journeys, often referred to as a "shift left" approach. That, in turn, would free up risk professionals to advise businesses on new product development and strategic business decisions, explore emerging risk trends and scenarios, strengthen resilience, and improve risk and control processes proactively.

These advances could lead to the creation of AI- and gen-AI-powered risk intelligence centers that serve all lines of defense (LODs): business and operations, the compliance and risk functions, and audits. Such a center would provide automated reporting, improved risk transparency, higher efficiency in risk-related decision making, and partial automation in drafting and updating policies and procedures to reflect changing regulatory requirements. It would act as a reliable and efficient source of information, enabling risk managers to make informed decisions swiftly and accurately.

For instance, McKinsey has developed a gen AI virtual expert that can provide tailored answers based on the firm's proprietary information and assets. Banks' risk functions and their stakeholders can develop similar tools that scan transactions with other banks, potential red flags, market news, asset prices, and more to influence risk decisions. These virtual experts can also collect data and evaluate climate risk assessments to answer counterparty questions.

Finally, gen AI could facilitate better coordination between the first and second LODs in the organization while maintaining the governance structure across all three. The improved coordination would enable enhanced monitoring and control mechanisms, thereby strengthening the organization's risk management framework.

Emerging applications of gen AI in risk and compliance

Of the many promising applications of gen AI for financial institutions, there's a set of candidates that banks are exploring for a first wave of adoption: regulatory compliance, financial crime, credit risk, modeling and data analytics, cyber risk, and climate risk. Overall, we see applications of gen AI across risk and compliance functions through three use case archetypes.

Through a *virtual expert*, a user can ask a question and receive a generated summary answer that's built from long-form documents and unstructured data. With *manual process automation*, gen AI performs time-consuming tasks. With *code acceleration*, gen AI updates or translates old code or writes entirely new code. All these archetypes can have roles in the key responsibilities of risk and compliance:

— *Regulatory compliance.* Enterprises are using gen AI as a virtual regulatory and policy expert by training it to answer questions about regulations, company policies, and guidelines. The tech can also compare policies, regulations, and operating procedures. As a code accelerator, it can check code for compliance misalignment and gaps. It can automate checking of regulatory compliance and provide alerts for potential breaches.

— *Financial crime.* Gen AI can generate suspicious-activity reports based on customer and transaction information. It can also automate the creation and update of customers' risk ratings based on changes in know-your-customer attributes. By generating and improving code to detect suspicious activity and analyze transactions, the tech can improve transaction monitoring.

— *Credit risk.* By summarizing customer information (for example, transactions with other banks) to inform credit decisions, gen AI can help accelerate banks' end-to-end credit process. Following a credit decision, it can draft the credit memo and contract. Financial institutions are using the tech to generate credit risk reports and extract customer insights from credit memos. Gen AI can generate code to source and analyze credit data to gain a view into customers' risk profiles and generate default and loss probability estimates through models.

— *Modeling and data analytics.* Gen AI can accelerate the migration of legacy programming languages, such as the switch from SAS and COBOL to Python. It can also automate the monitoring of model performance and generate alerts if metrics fall outside tolerance levels. Companies are also using gen AI to draft model documentation and validation reports.

— *Cyber risk.* By checking cybersecurity vulnerabilities, gen AI can use natural language to generate code for detection rules and accelerate secure code development. It can be useful in "red teaming" (simulating adversarial strategies and testing attack scenarios). The tech can also serve as a virtual expert for investigating security data. It can make risk detection smarter by speeding and aggregating security insights and trends from security events and behavior anomalies.

— *Climate risk.* As a code accelerator, gen AI can suggest code snippets, facilitate unit testing, and assist physical-risk visualization with high-resolution maps. It can automate data collection for counterparty transition risk assessments and generate early-warning signals based on trigger events. As a virtual expert, gen AI can automatically generate reports on environmental, social, and governance (ESG) topics and sustainability sections of annual reports (see sidebar, "How generative AI can speed financial institutions' climate risk assessments").

Once companies have embedded gen AI in these roles and functions, they have seen a second wave of emerging use cases across other aspects of risk management. Gen AI can streamline enterprise risk by synthesizing enterprise-risk-management summaries from existing data and reports. It can help accelerate the internal capital adequacy assessment process and model capital adequacy by sourcing relevant data. Banks can also use it to summarize risk positions and draft risk reports and executive briefings for senior management.

Another area in which gen AI can play an important role is operational risk. Banks can use it for operational automation of controls, monitoring, and incident detection. It can also automatically draft risk and control self-assessments or evaluate existing ones for quality.

Key considerations in gen AI adoption

While several compelling use cases exist in which gen AI can propel productivity, prioritizing them is critical to realizing value while adopting the tech responsibly and sustainably. We see three critical dimensions that risk leaders can assess to determine prioritization of use cases and maximize impact (exhibit).

Chief risk officers can base their decisions on assessments across qualitative and quantitative dimensions of impact, risk, and feasibility. This process includes aligning with their banks' overall visions for gen AI and associated guardrails, understanding relevant regulations (such as the EU AI Act), and assessing data sensitivity. All leaders need to be aware of the novel risks associated with this new tech. These risks can be broadly divided into eight categories:

How generative AI can speed financial institutions' climate risk assessments

Risk functions can benefit from generative AI (gen AI) across a variety of analyses. In the case of climate risk assessments, the technology—via tools based on generative pretrained transformers—can instantaneously draw from multiple, lengthy reports and distill answers from source materials (exhibit).

In addition, gen AI can provide support to relationship managers to accelerate the assessment of climate risk for their counterparties. It can automatically generate syntheses of counterparty transition plans and compare them against actual emissions to evaluate progress toward goals.

Beyond measurement, gen AI can aid climate impact analysis by ultimately automating reporting on environmental, social, and governance topics. It can aid risk by automating climate risk drafts, and it can spur growth by using customer data to personalize green financial products.

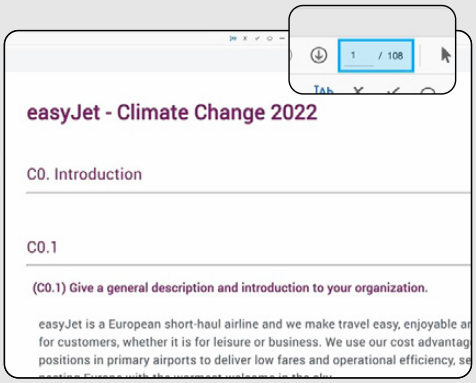
Consider the benefits of gen AI automation in helping customers move to net zero. The tech can identify market trends and environmental impact from years of company reports. In turn, financial institutions can use that new information to find investment opportunities.

Exhibit

How a virtual expert can be used to accelerate climate risk assessments.

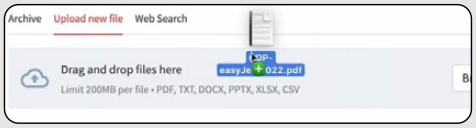
Generative AI (gen AI) analyzes and processes data to speed climate reporting (illustrative)

1



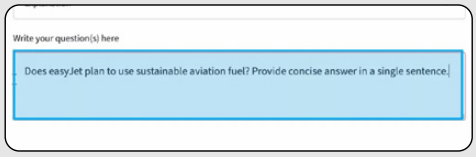
The GPT-based insight extraction engine extracts information from 100+ page sustainability reports

2

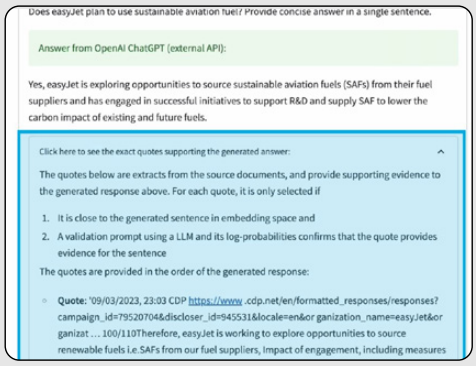


The gen AI solution preprocesses files and identifies relevant paragraphs for frontline bankers to quickly find insights about their clients' sustainability plans

3



4



The gen AI solution synthesizes answers drawing from many sources, extracts supporting quotes, and gives confidence levels to answers

Source: CDP report 2023, formerly Carbon Disclosure Project

McKinsey & Company

— impaired fairness, when the output of a gen AI model may be inherently biased against a particular group of users

— intellectual property infringement, such as copyright violations and plagiarism incidents, as foundation models typically leverage internet-based data

— privacy concerns, such as unauthorized public disclosure of personal or sensitive information

— malicious use, such as dissemination of false content and use of gen AI by criminals to create false identities, orchestrate phishing attacks, or scam customers

— security threats, when vulnerabilities within gen AI systems can be breached or exploited

— performance and "explainability" risks, such as models providing factually incorrect answers and outdated information

— strategic risks through noncompliance with ESG standards or regulations, creating societal or reputational risks

— third-party risks, such as leakage of proprietary data to the public realm through the use of third-party tools

Winning strategies for planning a gen AI journey

Organizations that can extract value from gen AI should use a focused, top-down approach to start the journey. Given the scarcity of talent to scale gen AI capabilities, organizations should start with

Exhibit

Risk leaders can prioritize risk, impact, and feasibility considerations when planning gen AI implementation in a risk function.

Initial assessment to prioritize generative AI (gen AI) use cases based on impact, feasibility, and risk scoring[1]

Risk

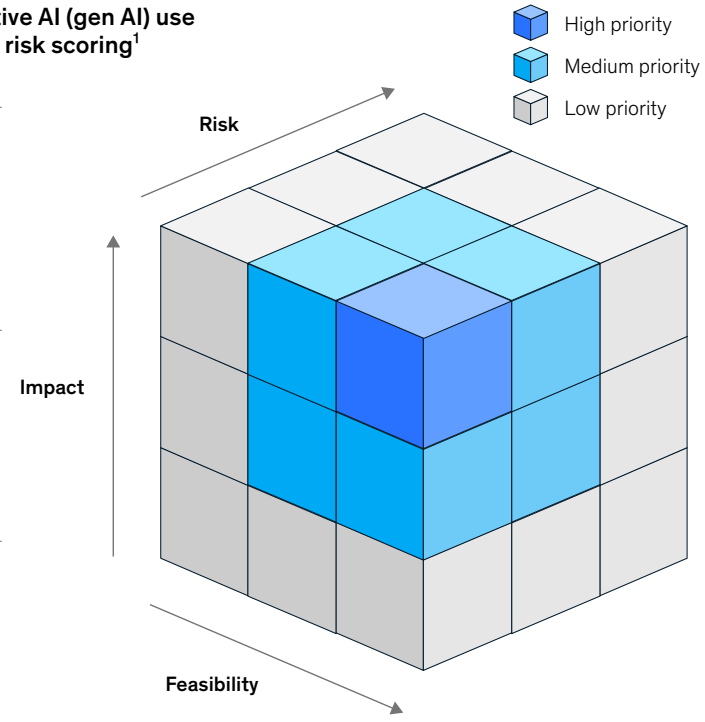
Source and use of data
Security threats
Performance and transparency
Strategic risk
Third-party risk

Impact

Revenue generation
Operational cost savings
Strategic priority of the organization
Scalability of gen AI ecosystem
Solves pain points not addressed by traditional AI

Feasibility

Data quality and architecture
Readiness of tech stack
Upskilling and hiring needs
Change management needs



[1]Taking quantitative and qualitative dimensions into account.

McKinsey & Company

three to five high-priority risk and compliance use cases that align with their strategic priorities. They can execute these use cases in three to six months, followed by an estimation of business impact. Scaling the applications will require the development of a gen AI ecosystem that focuses on seven areas:

- a catalog of production-ready, reusable gen AI services and solutions (use cases) that can be easily plugged into a range of business scenarios and applications across the banking value chain
- a secure, gen-AI-ready tech stack that supports hybrid-cloud deployments to enable support for unstructured data, vector embedding, machine learning training, execution, and pre- and postlaunch processing
- integration with enterprise-grade foundation models and tools to enable fit-for-purpose selection and orchestration across open and proprietary models
- automation of supporting tools, including MLOps (machine learning operations), data, and processing pipelines, to accelerate the development, release, and maintenance of gen AI solutions
- governance and talent models that readily deploy cross-functional expertise empowered to collaborate and exchange knowledge (such as language, natural-language processing, and reinforcement learning from human feedback, prompt engineers, cloud experts, AI product leaders, and legal and regulatory experts)
- process alignment for building gen AI to support the rapid and safe end-to-end experimentation, validation, and deployment of solutions
- a road map detailing the timeline for when various capabilities and solutions will be launched and scaled that aligns with the organization's broader business strategy

At a time when companies in all sectors are experimenting with gen AI, organizations that fail to harness the tech's potential are risking falling behind in efficiency, creativity, and customer engagement. At the outset, banks should keep in mind that the move from pilot to production takes significantly longer for gen AI than for classical AI and machine learning. In selecting use cases, risk and compliance functions may be tempted to use a siloed approach. Instead, they should align with an entire organization's gen AI strategy and goals.

For gen AI adoption by risk and compliance groups to be effective and responsible, it is critical that these groups understand the need for new risk management and controls, the importance of data and tech demands, and the new talent and operating-model requirements.

Risk management and controls

With gen AI, a new level of risk management and control is necessary. Winning responsibly requires both defensive and offensive strategies. All organizations face inbound risks from gen AI, in addition to the risks from developing gen AI use cases and embedding gen AI into standard workplace tools. So banks will need to evolve their risk mitigation capabilities accordingly.

The first wave heavily focuses on human-in-the-loop reviews to ensure the accuracy of model responses. Using gen AI to check itself, such as through source citations and risk scores, can make human reviews more efficient. By moving gen AI guardrails to real time and doing away with human-in-the-loop reviews, some companies are already putting gen AI directly in front of their customers. To make this move, risk and compliance professionals can work with development team members to set the guardrails and create controls from the start.

Risk functions need to be vigilant to manage gen AI risks at the enterprise level. They can fulfill that obligation by taking the following steps:

1. Ensure that everyone across the organization is aware of the risks inherent in gen AI, publishing dos and don'ts and setting risk guardrails.

2. Update model identification criteria and model risk policy (in line with regulations such as the EU AI Act) to enable the identification and classification of gen AI models, and have an appropriate risk assessment and control framework in place.

3. Develop gen AI risk and compliance experts who can work directly with frontline development teams on new products and customer journeys.

4. Revisit existing know-your-customer, anti–money laundering, fraud, and cyber controls to ensure that they are still effective in a gen-AI-enabled world.

Data and tech demands

Banks shouldn't underestimate the data and tech demands related to a gen AI system, which requires enormous amounts of both. Why? For one, the process of context embedding is crucial to ensure the accuracy and relevance of results. That process requires the input of appropriate data and addressing data quality issues. Moreover, the data on hand may be insufficient. Organizations may need to build or invest in labeled data sets to quantify, measure, and track the performance of gen AI applications based on task and use.

Data will serve as a competitive advantage in extracting value from gen AI. An organization looking to automate customer engagement using gen AI must have up-to-date, accurate data. Organizations with advanced data platforms will be the most effective at harnessing gen AI capabilities.

Talent and operating-model requirements

Since gen AI is a transformational technology requiring an organizational shift, organizations will need to understand the related talent requirements. Banks can embed operating-model changes into their culture and business-as-usual processes. They can train new users not only on how to use gen AI but also on its limitations and strengths. Assembling a team of "gen AI champions" can help shape, build, and scale adoption of this new tech.

We expect gen AI to empower banks' entire risk and compliance functions in the future. This implies a profound culture change that will require all risk professionals to be conversant with the new tech, its capabilities, its limitations, and how to mitigate those limitations. Using gen AI will be a significant shift for all organizations, but those that navigate the delicate balance of harnessing the technology's powers while managing the risks it poses can achieve significant productivity gains.

Rahul Agarwal is an associate partner in McKinsey's New Jersey office, **Andreas Kremer** is a partner in the Berlin office, **Ida Kristensen** is a senior partner in the New York office, and **Angela Luget** is a partner in the London office.

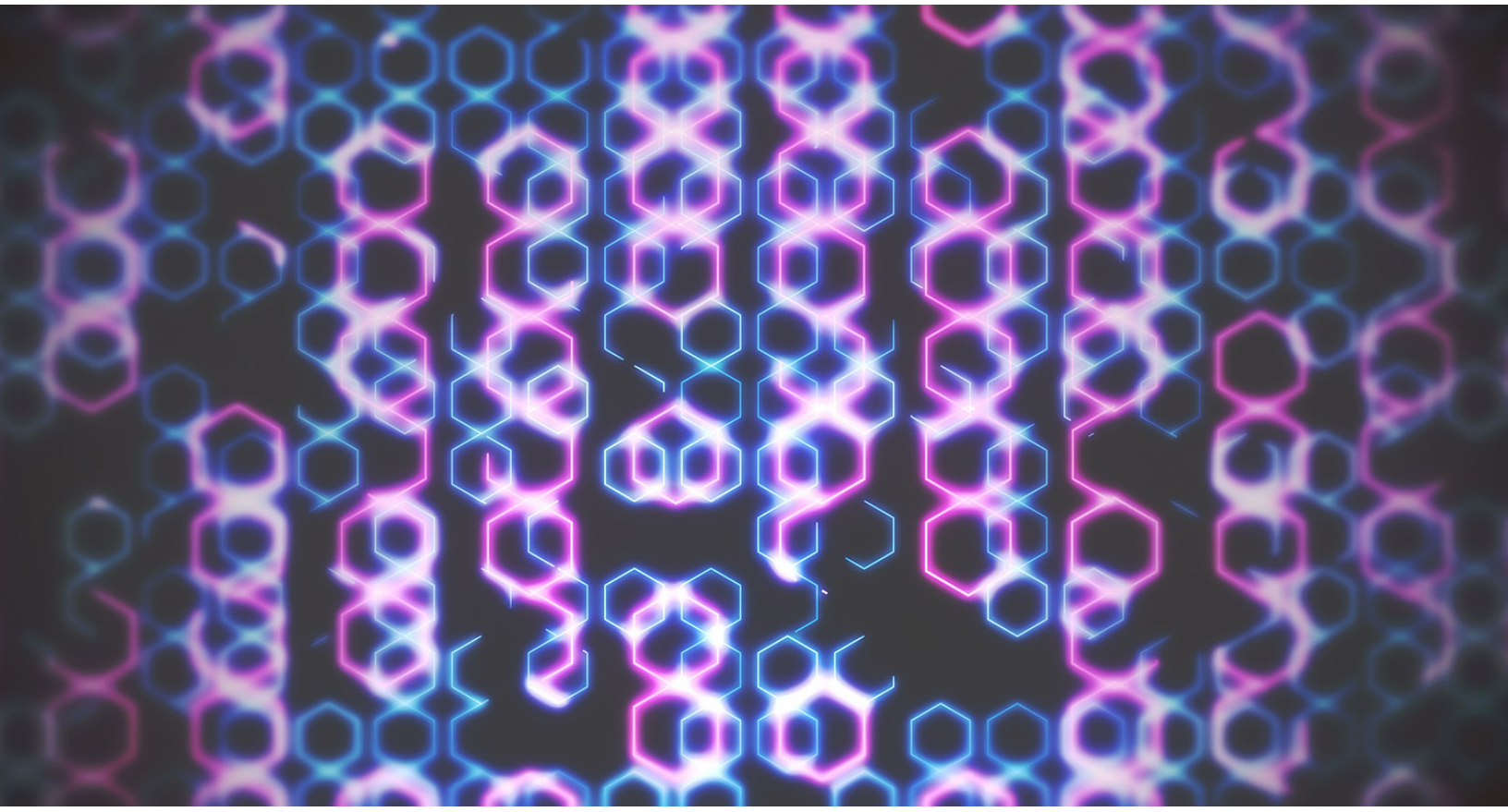
The authors wish to thank Adrija Banerjee, Stephan Beitz, Adrian Foerster, Yilin Li, Anke Raufuss, Ibtesam Siddiqui, and Claudia Satrústegui for their contributions to this article.

Copyright © 2024 McKinsey & Company. All rights reserved.

As gen AI advances, regulators—and risk functions—rush to keep pace

AI and its supercharged breakthrough, generative AI, are all about rapid advancements, and rule makers are under pressure to keep up.

This article is a collaborative effort by Andreas Kremer, Angela Luget, Daniel Mikkelsen, Henning Soller, Malin Strandell-Jansson, and Sheila Zingg, representing views from McKinsey's Risk & Resilience Practice.



© Getty Images

The rapid advancement of generative AI (gen AI) has regulators around the world racing to understand, control, and guarantee the safety of the technology—all while preserving its potential benefits. Across industries, gen AI adoption has presented a new challenge for risk and compliance functions: how to balance use of this new technology amid an evolving—and uneven—regulatory framework.

As governments and regulators try to define what such a control environment should look like, the developing approaches are fragmented and often misaligned, making it difficult for organizations to navigate and causing substantial uncertainty.

In this article, we explain the risks of AI and gen AI and why the technology has drawn regulatory scrutiny. We also offer a strategic road map to help risk functions navigate the uneven and changing rule-making landscape—which is focused not only on gen AI but all artificial intelligence.

Why does gen AI need regulation?

AI's breakthrough advancement, gen AI, has quickly captured the interest of the public, with ChatGPT becoming one of the fastest-growing platforms ever, reaching one million users in just five days. The acceleration comes as no surprise given the wide range of gen AI use cases, which promise increased productivity, expedited access to knowledge, and an expected total economic impact of $2.6 trillion to $4.4 trillion annually.[1]

There is, however, an economic incentive to getting AI and gen AI adoption right. Companies developing these systems may face consequences if the platforms they develop are not sufficiently polished. And a misstep can be costly. Major gen AI companies, for example, have lost significant market value when their platforms were found hallucinating (when AI generates false or illogical information).

The proliferation of gen AI has increased the visibility of risks. Key gen AI concerns include how the technology's models and systems are developed and how the technology is used.

Generally, there are concerns about a potential lack of transparency in the functioning of gen AI systems, the data used to train them, issues of bias and fairness, potential intellectual property infringements, possible privacy violations, third-party risk, as well as security concerns.

Add disinformation to these concerns, such as erroneous or manipulated output and harmful or malicious content, and it is no wonder regulators are seeking to mitigate potential harms. Regulators seek to establish legal certainty for companies engaged in the development or use of gen AI. Meanwhile, rule makers want to encourage innovation without fear of unknown repercussions.

The goal is to establish harmonized international regulatory standards that would stimulate international trade and data transfers. In pursuit of this goal, a consensus has been reached: the gen AI development community has been at the forefront of advocating for some regulatory control over the technology's development as soon as possible. The question at hand is not whether to proceed with regulations, but rather how.

The current international regulatory landscape for AI

While no country has passed comprehensive AI or gen AI regulation to date, leading legislative efforts include those in Brazil, China, the European Union, Singapore, South Korea, and the United States. The approaches taken by the different countries vary from broad AI regulation supported by existing data protection and cybersecurity regulations (the European Union and South Korea) to sector-specific laws (the United States) and more general principles or guidelines-based approaches (Brazil, Singapore, and the United States). Each approach has its own benefits and drawbacks, and some markets will move from principles-based guidelines to strict legislation over time (Exhibit 1).

While the approaches vary, common themes in the regulatory landscape have emerged globally:

— *Transparency.* Regulators are seeking traceability and clarity of AI output. Their goal

[1] "The economic potential of generative AI: The next productivity frontier," McKinsey, June 14, 2023.

is to ensure that users are informed when they engage with any AI system and to provide them with information about their rights and about the capabilities and limitations of the system.

— *Human agency and oversight.* Ideally, AI systems should be developed and used as tools that serve people, uphold human dignity and personal autonomy, and function in a way that can be appropriately controlled and overseen by humans.

— *Accountability.* Regulators want to see mechanisms that ensure awareness of responsibilities, accountability, and potential redress regarding AI systems. In practice, they are seeking top management buy-in, organization-wide education, and awareness of individual responsibility.

— *Technical robustness and safety.* Rule makers are seeking to minimize unintended and unexpected harm by ensuring that AI systems are robust, meaning they operate as expected, remain stable, and can rectify user errors. They should have fallback solutions and remediation to address any failures to meet these criteria, and they should be resilient against attempts to manipulate the system by malicious third parties.

— *Diversity, nondiscrimination, and fairness.* Another goal for regulators is to ensure that AI systems are free of bias and that the output does not result in discrimination or unfair treatment of people.

— *Privacy and data governance.* Regulators want to see development and usage of AI systems that follow existing privacy and data protection rules while processing data that meet high standards in quality and integrity.

— *Social and environmental well-being.* There is a strong desire to ensure that all AI is sustainable, environmentally friendly (for instance, in its energy use), and beneficial to all people, with ongoing monitoring and assessing of the long-term effects on individuals, society, and democracy.

Despite some commonality in the guiding principles of AI, the implementation and exact wording vary by regulator and region. Many rules are still new and, thus, prone to frequent updates (Exhibit 2). This makes it challenging for organizations to navigate regulations while planning long-term AI strategies.

What does this mean for organizations?

Organizations may be tempted to wait to see what AI regulations emerge. But the time to act is now. Organizations may face large legal, reputational, organizational, and financial risks if they do not act swiftly. Several markets, including Italy, have already banned ChatGPT because of privacy concerns,

Exhibit 1

Regulations related to AI governance vary around the world.

As of November 2023, nonexhaustive

Source: OECD; McKinsey analysis

McKinsey & Company

copyright infringement lawsuits brought by multiple organizations and individuals, and defamation lawsuits.

More speed bumps are likely. As the negative effects of AI become more widely known and publicized, public concerns increase. This, in turn, has led to public distrust of the companies creating or using AI.

A misstep at this stage could also be costly. Organizations could face fines from legal enforcement—of up to 7 percent of annual global revenues, according to the AI regulation proposed by the European Union, for example. Another threat is financial loss from falloff in customer or investor trust that could translate into a lower stock price, loss of customers, or slower customer acquisition. The incentive to move fast is heightened by the fact that if the right governance and organizational models for AI are not built early, remediation may become necessary later due to regulatory changes, data breaches, or cybersecurity incidents. Fixing a system after the fact can be both expensive and difficult to implement consistently across the organization.

The exact future of legal obligations is still unclear and may differ across geographies and depend on the specific role AI will play within the value chain. Still, there are some no-regret moves for organizations, which can be implemented today to get ahead of looming legal changes.

These preemptive actions can be grouped into four key areas that stem from existing data protection or privacy and cyber efforts, as they share a great deal of common ground:

Transparency. Create a taxonomy and inventory of models, classifying them in accordance with regulation, and record all usage across the organization in a central repository that is clear to those inside and outside the organization. Create detailed documentation of AI and gen AI usage, both internally and externally, its functioning, risks, and controls, and create clear documentation on how a model was developed, what risks it may have, and how it is intended to be used.

Governance. Implement a governance structure for AI and gen AI that ensures sufficient oversight, authority, and accountability both within the organization and with third parties and regulators. This approach should include a definition of all roles and responsibilities in AI and gen AI management and the development of an incident management plan to address any issues that may arise from AI and gen AI use. The governance structure should be robust enough to withstand changes in personnel and time but also agile enough to adapt to evolving technology, business priorities, and regulatory requirements.

Data, model, and technology management. AI and gen AI both require robust data, model, and technology management:

— *Data management.* Data is the foundation of all AI and gen AI models. The quality of the data input also mirrors the final output of the model. Proper and reliable data management includes awareness of data sources, data classification, data quality and lineage, intellectual property, and privacy management.

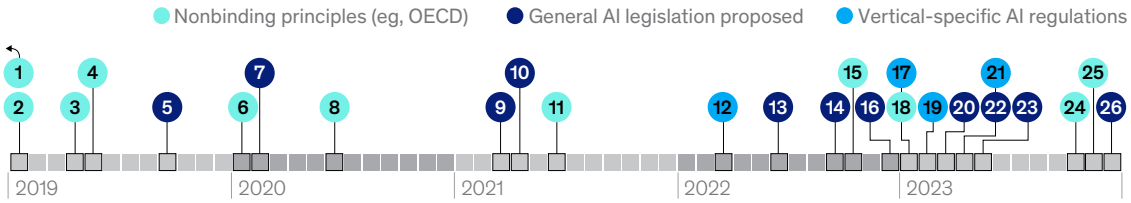
— *Model management.* Organizations can establish robust principles and guardrails for AI and gen AI development and use them to minimize the organization's risks and ensure that all AI and gen AI models uphold fairness and bias controls, proper functioning, transparency, clarity, and enablement of human oversight. Train the entire organization on the proper use and development of AI and gen AI to ensure risks are minimized. Develop the organization's risk taxonomy and risk framework to include the risks associated with gen AI. Establish roles and responsibilities in risk management and establish risk assessments and controls, with proper testing and monitoring mechanisms to monitor and resolve AI and gen AI risks. Both data and model management require agile and iterative processes and should not be treated as simple tick-the-box exercises at the beginning of development projects.

— *Cybersecurity and technology management.* Establish strong cybersecurity and technology, including access control, firewalls, logs, monitoring, etcetera, to ensure a secure technology environment, where unauthorized access or misuse is prevented and potential incidents are identified early.

Exhibit 2

AI governance–related policy and regulatory efforts are under way globally.

Examples by type of policy or effort and when proposed; nonexhaustive



2019 and earlier

- **1. Sept 2017** South Korea Ethical Guidelines for Intelligent Information Technology
- **2. Jan 2019** Singapore Model AI Governance Framework, first edition
- **3. Apr 2019** EU Ethics Guidelines for Trustworthy AI
- **4. May 2019** OECD AI Principles
- **5. Sept 2019** Bill establishing the principles for the use of AI in Brazil

2020

- **6. Jan 2020** Singapore Model AI Governance Framework, second edition
- **7. Feb 2020** Bill establishing the fundamental principles and guidelines for the development and application of AI in Brazil
- **8. June 2020** South Korea Framework Act on Intelligent Informatization

2021

- **9. Mar 2021** Bill providing for the ethical framework and guidelines that underlie the development and use of AI in Brazil
- **10. Apr 2021** Proposed EU AI Act (expires Q1 2024)
- **11. June 2021** South Korea Enforcement decree on Framework Act on Intelligent Informatization

2022

- **12. Mar 2022** China issues provisions on Internet Information Service Algorithm Recommendations and Administration of Deep Synthesis of Internet Information Services
- **13. June 2022** Canada's proposed Artificial Intelligence and Data Act (planned 2025)
- **14. Sept 2022** EU AI Liability Directive, a regime for dealing with damages caused by AI
- **15. Oct 2022** US Blueprint for an AI Bill of Rights
- **16. Dec 2022** Senate approval of the draft regulatory framework on artificial intelligence in Brazil

2023

- **17. Jan 2023** Stable Diffusion and Midjourney copyright lawsuits in the US
- **18. Jan 2023** NIST AI risk management framework
- **19. Feb 2023** South Korean Assembly proposed Act on Promotion of AI Industry and Framework for Establishing Trustworthy AI
- **20. Mar 2023** ChatGPT temporarily banned in Italy because of privacy concerns
- **21. Mar–Apr 2023** Several data protection regulators globally looking into ChatGPT data protection practices, eg, Germany, France, and Spain
- **22. Apr 2023** China released Draft Administrative Measures for Generative Artificial Intelligence Services
- **23. May 2023** Proposal for legal framework for artificial intelligence in Brazil merging previous proposals from 2019–21
- **24. Oct 2023** US presidential executive order on AI
- **25. Nov 2023** AI summit in UK
- **26. Dec 2023** Political agreement on EU AI Act

Source: OECD; McKinsey analysis

McKinsey & Company

Individual rights. Educate users: make them aware that they are interacting with an AI system, and provide clear instructions for use. This should include establishing a point of contact that provides transparency and enables users to exercise their rights, such as how to access data, how models work, and how to opt out. Finally, take a customer-centric approach to designing and using AI, one that considers the ethical implications of the data used and its potential impact on customers. Since not everything legal is necessarily ethical, it is important to prioritize the ethical considerations of AI usage.

AI and gen AI will continue to have a significant impact on many organizations, whether they are providers of AI models or users of AI systems. Despite the rapidly changing regulatory landscape, which is not yet aligned across geographies and sectors and may feel unpredictable, there are tangible benefits for organizations that improve how they provide and use AI now.

Failure to handle AI and gen AI prudently can lead to legal, reputational, organizational, and financial damages; however, organizations can prepare themselves by focusing on transparency, governance, technology and data management, and individual rights. Addressing these areas will create a solid basis for future data governance and risk reduction and help streamline operations across cybersecurity, data management and protection, and responsible AI. Perhaps more important, adopting safeguards will help position the organization as a trusted provider.

Andreas Kremer is a partner in McKinsey's Berlin office; **Angela Luget** is a partner in the London office, where **Daniel Mikkelsen** is a senior partner; **Henning Soller** is a partner in the Franfurt office; **Malin Strandell-Jansson** is a senior knowledge expert in the Stockholm office; and **Sheila Zingg** is a consultant in the Zurich office.

The authors wish to thank Rachel Lee, Chris Schmitz, and Angie Selzer for their contributions to this article.

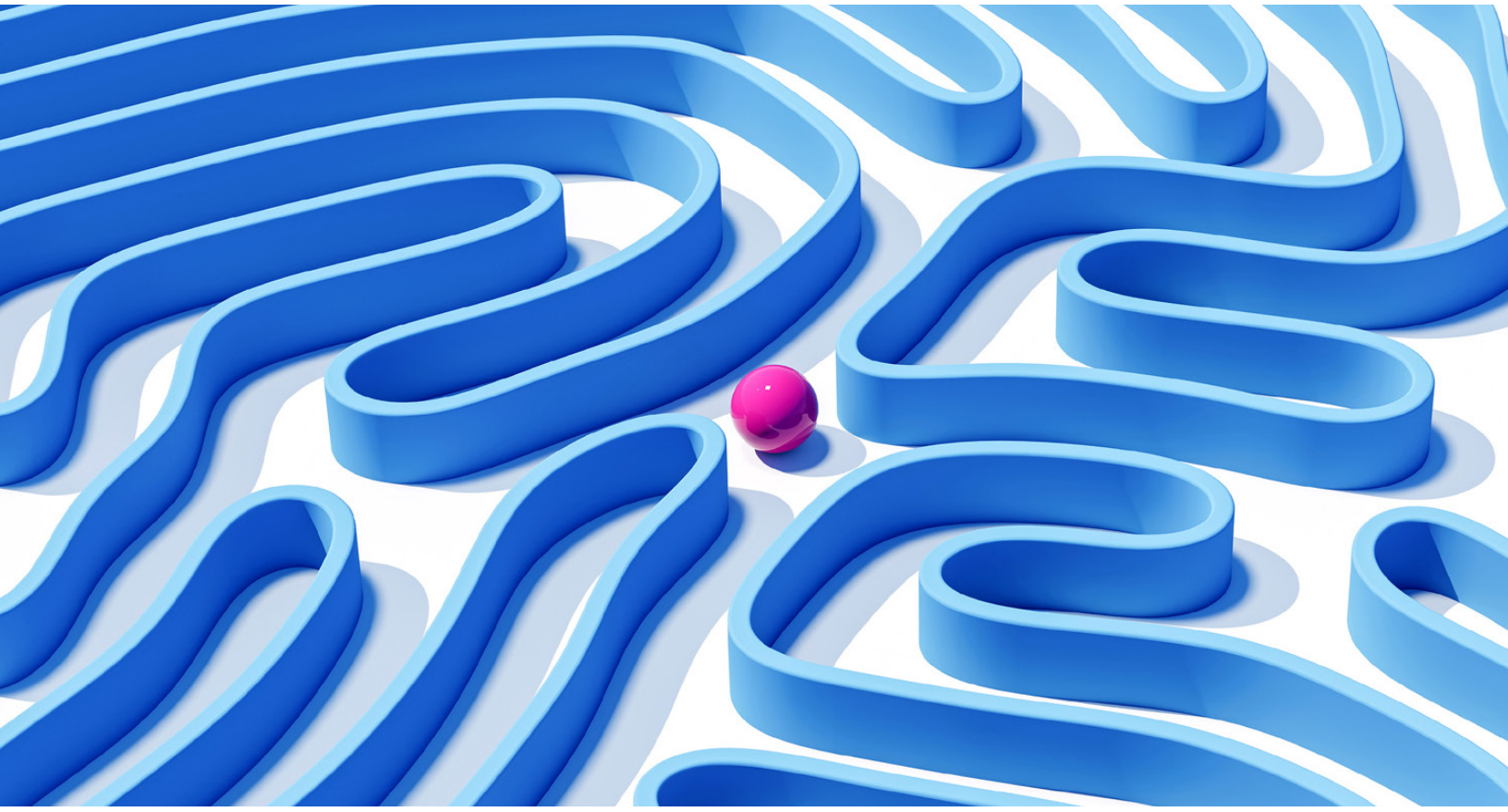
This article was edited by David Weidner, a senior editor in the Bay Area office.

Copyright © 2024 McKinsey & Company. All rights reserved.

How CEOs can mitigate compounding risks

When risks combine, the cumulative impact can have existential consequences. But leaders can prevent compounding risks from sneaking up on them by adapting risk processes to manage multiple threats.

by Ram Charan, Celia Huber, and Ophelia Usher



© Getty Images

One of a CEO's most important responsibilities is to create enduring value for shareholders. However, history shows how elusive that ambition is. Only 15 percent of the companies on the Fortune 500 list 50 years ago are still there today. Many once-iconic businesses ended up shutting down or being acquired because their leaders failed to address risks that they deemed insignificant, unlikely, far off in the future—or ones they didn't see at all.

In today's complex business environment, corporations face webs of intersecting risks whose combined impact is difficult to predict and manage. When several such hazards materialize simultaneously, the cumulative effect can pose an existential threat to the organization. Such compounding risks are particularly dangerous because management teams tend to underprepare for their combined impact. While corporate risk management processes track and strive to mitigate individual threats to the organization, they rarely assess the repercussions of several shocks occurring at once.

When a company's compounding risks turn into a full-blown crisis, industry peers, regulators, and commentators inevitably speculate about how the organization's management could have failed to address the looming threat. How, for example, could photography equipment makers have missed the smartphone revolution that ravaged their business when they created the first digital cameras? In most cases, the cause isn't willful ignorance or negligence but rather insufficient foresight: failing to ensure that the organization identifies potential compounding risks or delaying adequate actions to mitigate their impact. As complex, far-reaching risks mount, from geopolitical tensions to climate change, CEOs and boards cannot afford to be caught unprepared.

Three types of compounding risks

The threat of compounding risks has grown more severe because of the highly interconnected world that corporations operate in today. Often the causes of compounding risks are viewed as black swans, or "unknown unknowns" that no one could have foreseen, but in most cases the underlying risks can be predicted. To recognize them early on, leaders need to ask which known threats—from cyberattacks to technological disruptions to public health crises—could come together to create a compounding risk that should be considered in their risk mitigation strategies.

Compounding risks share two common features: the characteristics of the compounding risk are distinct from the underlying risks, and the compounding risk often has a different likelihood or impact than the underlying one. Further, compounding risks fall into three distinct categories: connected, cumulative, and novel risks.

Connected risks are threats to the business from multiple sources that leaders perceive as unrelated but that are in fact linked within a broader interconnected system. A single event that disrupts one part of the system can ripple out to other parts. For example, the COVID-19 pandemic caused declines in both regional manufacturing capacity and worldwide container-shipping capacity—two risks few companies anticipated being triggered by the same event. Similarly, during the 2008 financial crisis, organizations found suppliers and customers in disparate geographies going out of business as the crisis's fallout reverberated across global markets. Most recently, Russia's invasion of Ukraine created connected compounding risks for some organizations, such as a higher cost of raw materials combining with the sudden loss of international consumer markets.

In each case, a single risk—a pandemic, an economic crisis, a regional war—could have been existential in its own right. Most leaders realize that such significant events would disrupt their businesses, but the way these crises ricocheted across an interconnected business world caught many by surprise.

The second category of compounding risks is *cumulative risks*, whereby one or more risks build over time to trigger a single major shock. The underlying risks are often known to management teams and may even be rigorously monitored. However, the metrics usually only track individual incidents (for example, how often an IT system goes down) and the thresholds for alerting senior management are set high (such

Risks stemming from multiple sources that management perceives as being unrelated but that are linked within a broader interconnected system.

Example: Russia's invasion of Ukraine led to higher costs of raw materials and the loss of international consumer markets.

Risks that may be minor but as they build over time, they trigger a single major shock.

Example: A few negative social media posts spread virally, eventually damaging an organization's reputation and causing a customer exodus.

Material risks that combine to create an unexpected new risk with distinct characteristics.

Example: Technological acceleration spurred by the COVID-19 pandemic, combined with cryptocurrency miners' high demand for microchips, created a worldwide chip shortage.

as a certain percentage of accounts being past due). As a result, leaders are often unaware that the frequency or severity of these risks is mounting. Just like compounding interest, they accumulate, exacerbating the threat as years pass in part because the second- and third-order consequences may not be considered. For example, the risk framework may estimate the percentage of transactions lost during a single IT outage but not the potential lifetime revenue of a lost customer or the reputational damage and possible customer exodus caused by repeated outages.

Author Malcolm Gladwell defines this tipping point as "the moment of critical mass, the threshold, the boiling point."[1] Just as a single bump in the road may not cause a loose tire to fall off but a longer rough stretch does, so an individual event may be manageable but a series of them can become an existential threat. One industrial company faced near bankruptcy when bad acquisitions, high debt, and a bloated balance sheet left it deeply exposed to the fallout of the 2008 financial crisis. Or consider catastrophic industrial accidents: thanks to modern safety protocols, a single failure point is unlikely to cause a disaster, but multiple safety failures occurring simultaneously can become a crisis. Social media is a frequent source of this type of compounding risk because a few negative tweets or posts can spread virally, perpetuating a (potentially false) narrative that deeply damages an organization's reputation.

The final form of compounding risk, which we call *novel risk*, involves multiple known material risks—be they cyberattacks or threats to the business model or vulnerabilities caused by financial maneuvers—combining to create an unexpected new risk with distinct characteristics. The underlying risks are often long-term in nature, such as the impact of climate change, geopolitical tensions, or technological disruptions. Recent years have provided ample illustration of the dangers that a sudden new risk layered onto existing risks can pose. Companies carrying large debt loads were able to manage that risk—until the pandemic battered their returns. Cryptocurrency miners' high demand for microchips seemed tangential to many businesses—until pandemic-induced technological acceleration and supply chain problems created a worldwide chip shortage that brought numerous manufacturers to a standstill.

In most cases, the underlying risks are on companies' radars. The novel challenges these risks could create in combination, however, are not.

[1] Malcolm Gladwell, *The Tipping Point: How Little Things Can Make a Big Difference*, Boston, MA: Little, Brown and Company, 2000.

Neither is the need for risk functions, management, and boards to pressure-test their ability to navigate such compounding risks.

How to address compounding risks

As the individual responsible for balancing the company's short-term performance with long-term prosperity, the CEO holds ultimate responsibility for addressing compounding risks. To get a handle on such threats, leaders can take four steps: ensure their risk governance program covers compounding risks, validate that their teams are adequately prepared to manage such risks, leverage a horizon approach to investing to ensure long-term vectors of compounding risk are not ignored, and consider compound-risk scenarios when planning big strategic bets.

Strengthen risk management governance. Leaders should instruct their risk management functions to broaden the aperture on the risk scenarios they monitor to include compounding risks. For example, once risk managers have identified the top risks to the business, they often create an enterprise-level risk management map. Instead, the team should consider how and which individual risks could combine to create a new compounding risk, with particular focus on risks that may be minor individually but have high frequency (IT outages, for instance). Looking at the business through the lens of the customer rather than through product offerings can help risk managers see small but recurring friction points that could cause customers to leave.

All risks are best tracked through a formal risk management process. It's critical to establish accountability, with senior executives' performance scorecards linked to risk management goals and boards regularly updated on how management is preparing for compounding risks (see sidebar, "The board's role in addressing compounding risk.") Establishing early warning signals will allow leaders to see how risks are evolving. For example, what leading indicators are you monitoring to understand how shifting or escalating geopolitical tensions could create compounding risk for your operations in the short or long term?

Run "premortems" on managing risks. The current volatility has led many organizations to embrace scenarios in strategic planning, but most "what if" constructs don't cover the full range of

The board's role in addressing compounding risk

While addressing risks is management's responsibility, the board should ensure that senior executives are considering and mitigating risks critical to the company's long-term performance.

- Ask what risk scenarios related to business, economic, and market conditions leadership has considered, and if it has dismissed them, probe why.
- Pressure-test the management team's risk assessment process, especially regarding compounding risks, potentially bringing in experts to help run through "what if" scenarios.
- Pressure-test leadership's plan to address risks. For example, how is the organization investing in different strategic horizons? Do the big bets leadership is making address compound risks, not just individual risks?
- Make sure leadership is tracking and sharing leading indicators that could alert the board to a changing business environment and the need to change strategies.
- Assess big capital investment plans against multiple "what if" scenarios to understand how those decisions create or address risks.
- Ensure the organization is financially resilient enough to withstand multiple shocks and verify the vulnerabilities of the existing business model to those shocks.
- Know your role in a crisis and prepare for it by running exercises on acute scenarios such as a massive cyberattack.

compounding risks. Analyzing factors that could produce a crisis can help management teams identify compounding risks and their consequences across multiple time horizons. In such premortem sessions, the team assumes a major negative event (for example, a 75 percent sales drop), then works backward to imagine how such a scenario might occur. What products could customers use as substitutes for your offerings? What could cause them to switch? What occurrence could critically harm the company's reputation?

During workshops or executive retreats, futurists and other experts from inside and outside the organization can help the leadership team recognize compounding risks they may otherwise not consider. The key to a successful premortem is having a "challenger" mindset and reviewing multiple scenarios in which compounding risks can lead to a crisis.

Use a horizon planning approach. Many compounding risks stem from trends with long-term time horizons such as climate change, market or business model innovations, or changing consumer behaviors. These risks tend to build slowly until they hit the tipping point of becoming existential for the organization. A horizon planning approach can help management teams address risks that can emerge at various stages by looking at three horizons: first, maintaining and defending the core business; second, nurturing emerging businesses; and third, creating genuinely new businesses.

Addressing the last horizon is particularly important to mitigating long-term risks. For example, many energy companies are investing in decarbonizing their businesses even as they continue to rely on fossil fuels. Likewise, most car companies are developing electric cars while continuing to sell gasoline-powered vehicles. In essence, the horizon approach prepares companies for the next industry disruption—which often takes the form of a compounding risk, such as a combination of regulatory changes, consumer behavior shifts, and technological advances.

Make big bets that address long-term risks. As part of the horizon approach, the CEO needs to make big strategic bets that can fundamentally change the organization's trajectory. Such investments enable a company to evolve along with its industry and in the process hedge long-term risks. However, these big bets should not be aimed at neutralizing a single risk but at mitigating numerous threats the organization faces, as industry disruptions are likely to stem from a confluence of risks.

Compounding risks are often missed by risk management functions, CEOs, and boards. Yet when unidentified, unmonitored, and unaddressed, they can threaten organizations' survival. With mounting geopolitical tensions, rapid technological shifts, and other long-term threats that have wide-ranging implications, CEOs need to ensure that their organizations are tracking the interactions among different risks and are prepared for multiple crises striking simultaneously.

Celia Huber is a senior partner in McKinsey's Bay Area office, and **Ophelia Usher** is an expert associate partner in the Stamford office. **Ram Charan** is an author and business consultant to CEOs and boards.

This article was edited by Joanna Pachner, an executive editor in the Toronto office.

Copyright © 2024 McKinsey & Company. All rights reserved.

McKinsey & Company | QuantumBlack AI by McKinsey

Implementing generative AI with speed and safety

Generative AI poses both risks and opportunities. Here's a road map to mitigate the former while moving to capture the latter from day one.

This article is a collaborative effort by Oliver Bevan, Michael Chui, Ida Kristensen, Brittany Presten, and Lareina Yee, representing views from McKinsey's Risk & Resilience Practice and QuantumBlack, AI by McKinsey.

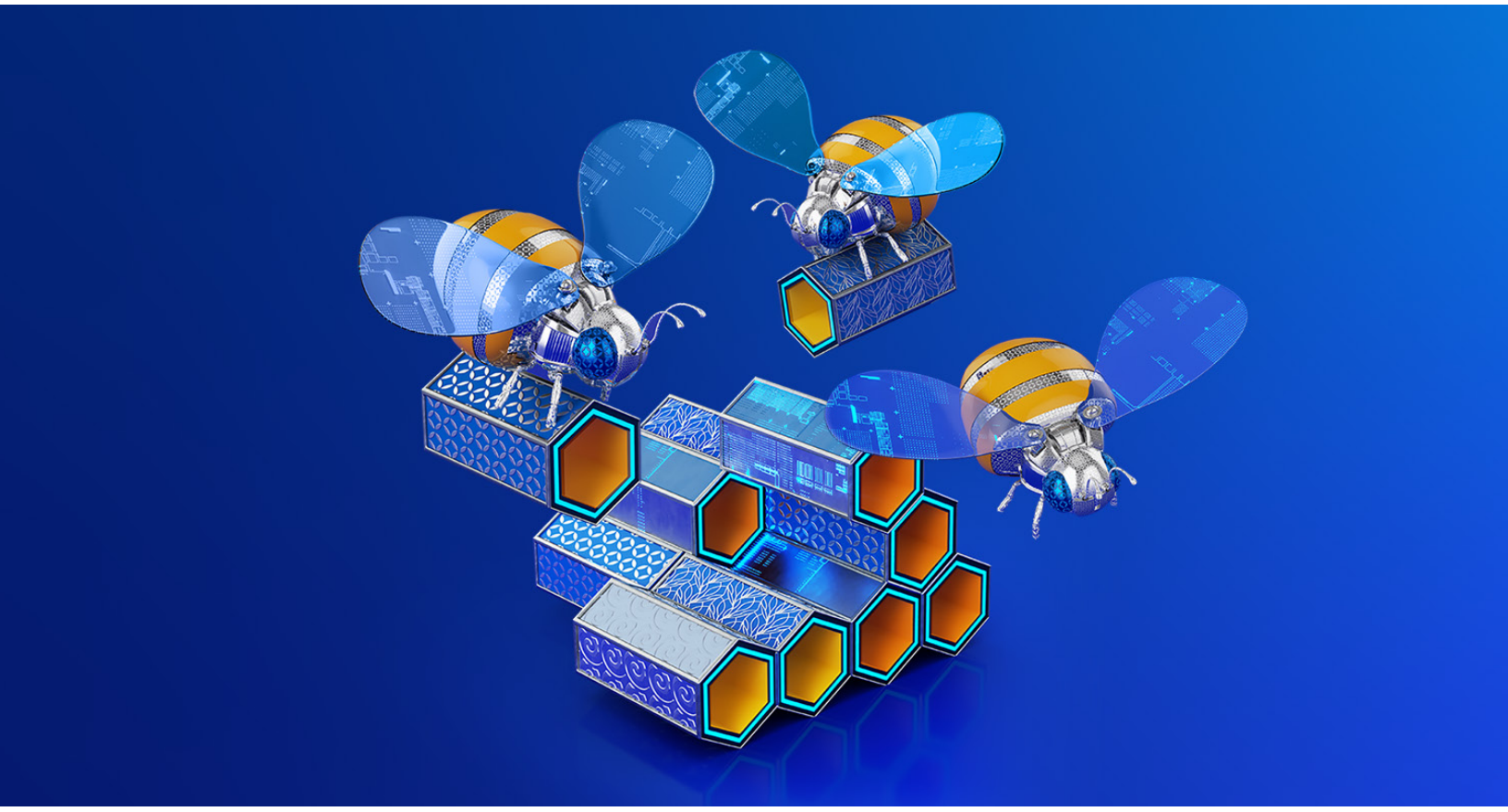


Illustration by Graham Hutchings

Generative AI (gen AI) presents a once-in-a-generation opportunity for companies, with the potential for transformative impact across innovation, growth, and productivity. The technology can now produce credible software code, text, speech, high-fidelity images, and interactive videos. It has identified the potential for millions of new materials through crystal structures and even developed molecular models that may serve as the base for finding cures for previously untreated diseases.

McKinsey research has estimated that gen AI has the potential to add up to $4.4 trillion in economic value to the global economy while enhancing the impact of all AI by 15 to 40 percent.[1] While many corporate leaders are determined to capture this value, there's a growing recognition that gen AI opportunities are accompanied by significant risks. In a recent flash survey of more than 100 organizations with more than $50 million in annual revenue, McKinsey finds that 63 percent of respondents characterize the implementation of gen AI as a "high" or "very high" priority.[2] Yet 91 percent of these respondents don't feel "very prepared" to do so in a responsible manner.

That unease is understandable. The risks associated with gen AI range from inaccurate outputs and biases embedded in the underlying training data to the potential for large-scale misinformation and malicious influence on politics and personal well-being. There are also broader debates on both the possibility and desirability of developing AI in general. These issues could undermine the judicious deployment of gen AI, potentially leading companies to pause experimentation until the risks are better understood—or even deprioritize the technology because of concerns over an inability to manage the novelty and complexity of these issues.

However, by adapting proven risk management approaches to gen AI, it's possible to move responsibly and with good pace to capture the value of the technology. Doing so will also allow companies to operate effectively while the regulatory environment around AI continues to evolve, such as with President Biden's executive order regarding gen AI development and use and the EU AI Act (see sidebar, "The United States moves to regulate AI"). In addition, most organizations are likely to see the use of gen AI increase "inbound" threats (risks likely to affect

The United States moves to regulate AI

On October 30, 2023, the Biden administration released a long-awaited executive order aimed at addressing concerns related to AI development in economic, national-security, and social domains. The order establishes principles, tasks federal agencies with AI-testing methods, codifies government oversight of private AI development, and outlines AI's impact on national security and foreign policy:

- *Holistic AI governance.* The order establishes a comprehensive framework for AI governance, emphasizing ethics, safety, and security. It addresses the importance of responsible innovation, collaboration, and competition in the AI industry.
- *Private sector accountability.* The order mandates that private companies involved in AI adhere to industry standards, report on compliance, and implement best practices. This includes meeting specific guidelines on transparency and accountability, especially for dual-use foundation models and large-scale computing clusters.
- *Cross-sector impact.* The order addresses various sectors affected by AI, including critical infrastructure, cybersecurity, education, healthcare, national security, and transportation. It promotes interagency collaboration to integrate AI responsibly and securely across these sectors, aligning government and industry efforts for societal benefit.

[1] "The economic potential of generative AI: The next productivity frontier," McKinsey, June 14, 2023.
[2] Unpublished data from McKinsey survey results.

organizations regardless of whether they deploy gen AI), particularly in fraud and cyber domains (early indications are that gen AI will be able to defeat standard antifraud biometric checks[3]). Building fit-for-purpose risk management will help guard against these threats.

In practical terms, enterprises looking to address gen AI risk should take the following four steps:

1. Launch a sprint to understand the risk of inbound exposures related to gen AI.

2. Develop a comprehensive view of the materiality of gen-AI-related risks across domains and use cases, and build a range of options (including both technical and nontechnical measures) to manage risks.

3. Establish a governance structure that balances expertise and oversight with an ability to support rapid decision making, adapting existing structures whenever possible.

4. Embed the governance structure in an operating model that draws on expertise across the organization and includes appropriate training for end users.

The specifics of how to implement these steps and the degree of change required to make them effective will vary with an organization's gen AI aspirations and nature. For instance, it could be looking to be a *maker* of the foundation models, a *shaper* that customizes and scales foundation models, or a *taker* that adopts foundation models through off-the-shelf applications with little or no customization (for example, standard office productivity software).[4]

This article provides a blueprint for developing an approach to implementing gen AI responsibly. Following these steps helps organizations move quickly to scale the technology and capture its benefits while minimizing their exposure to the potential downsides.

Understanding and responding to inbound risks

In our experience, including through building McKinsey's own gen AI application, gen-AI-related risks can be captured in eight main categories (Exhibit 1). These categories consider both inbound risks and risks that directly result from the adoption of gen AI tools and applications. Every company should develop some version of this core taxonomy to support understanding and communication on the risks arising from the implementation of gen AI.

Most organizations are likely to see the use of gen AI increase 'inbound' threats, particularly in fraud and cyber domains.

[3] *Security Intelligence*, "AI may soon defeat biometric security, even facial recognition software," blog entry by Mike Elgan, January 31, 2019.
[4] For more, see "Technology's generational moment with generative AI: A CIO and CTO guide," McKinsey, July 11, 2023.

Exhibit 1

Half of eight basic categories of generative AI risk apply to all organizations regardless of their deployment of related use cases.

Risk category	Description	Inbound	Gen AI[1] adoption
Impaired fairness	Algorithmic bias resulting from unrepresentative training data or model performance or misrepresentation of AI-generated content as human created		✓
Intellectual property (IP) infringement	Infringement on copyrighted or otherwise legally protected materials, inadvertent leakage of IP into public domain, or both	✓	✓
Data privacy and quality	Unauthorized use or disclosure of personal or sensitive information or use of incomplete or inaccurate data for model training		✓
Malicious use	Malicious or harmful AI-generated content (eg, falsehoods/deepfakes, scams/phishing, hate speech)	✓	✓
Security threats	Vulnerabilities in gen AI systems (eg, payload splitting to bypass safety filters, manipulability of open-source models)	✓	✓
Performance and "explainability"	Inability to explain model outputs or model inaccuracies appropriately (eg, factually incorrect or outdated answers, hallucinations)		✓
Strategic	Risk of noncompliance with standards or regulations, societal risk, and reputational risk		✓
Third party	Risks associated with use of third-party AI tools (eg, proprietary data being used by public models)	✓	✓

[1]Generative AI.

McKinsey & Company

Deciding how to respond to inbound risks is a focus for many executive teams and boards. This decision should serve as a foundation for how an organization communicates about gen AI to its employees and stakeholders. It should also inform the approach to use cases.

We see four primary sources of inbound risk from the adoption of gen AI:

- security threats, resulting from the increased volume and sophistication of attacks from gen-AI-enabled malware
- third-party risk, resulting from challenges in understanding where and how third parties may be deploying gen AI, creating potential unknown exposures
- malicious use, resulting from the potential for bad actors to create compelling deepfakes of company representatives or branding that result in significant reputational damage
- intellectual property (IP) infringement, resulting from IP (such as images, music, and text) being scraped into training engines for underlying large language models and made accessible to anyone using the technology

Most organizations will benefit from a focused sprint to investigate how gen AI is changing their external environment, with two primary objectives. The first is to understand potential exposures to inbound risks, anchored in the organization's risk profile (for example, how many third parties have access to sensitive or confidential data that need to be restricted from training external gen AI models). The second objective is to understand the maturity and readiness of the control environment—the technical and nontechnical capabilities the organization has in place to prevent, detect, and ultimately respond to inbound risks. These include cyber and fraud defenses, third-party diligence to identify where critical third parties may be deploying gen AI, and the ability to limit the scraping of company IP by engines used to train large language models.

The outcome of these efforts should be an understanding of where the organization faces the largest potential inbound exposures, as well as the maturity and readiness of its current defense system. Having conducted this exercise, the organization should have a clear road map of where to harden defenses and what the potential ROI from these efforts would be in potential risk mitigation.

Given the evolving nature of the technology underlying gen AI and its applications, organizations will need to repeat the effort to identify their exposure with some regularity. For most organizations, refreshing this exercise at least semiannually will be important until the pace of change has moderated and the control environments and defenses have matured.

Tethering Prometheus: Managing the risks produced by gen AI adoption

Organizations with ambitions to deploy gen AI will need to undertake additional, ongoing efforts to understand and manage the risks of the technology's adoption. This will likely require an investment of time and resources and a shift in ways of working. Yet it's essential if organizations are to achieve long-term, sustainable, and transformative benefits from gen AI. Missteps and failures can erode the confidence of executives, employees, and customers and trigger scaling back in the level of ambition to ultrasafe use cases that generate limited risk but are also unlikely to capitalize on the technology's true potential.

Organizations looking to deploy high-potential use cases for gen AI to drive productivity and innovation; provide better, more consistent customer service; and boost creativity in marketing and sales must address the challenge of responsible implementation. These use cases have varying risk profiles, reflecting both the nature of the technology itself and company-specific context concerning the specifics of the use case (for example, deployment of a gen AI chatbot to certain at-risk populations has a very different risk profile from that of a B2B deployment) (Exhibit 2).

Exhibit 2

Different generative AI use cases are associated with different kinds of risk.

Primary risk

Generative AI use case	Impaired fairness	IP[1] infringement	Data privacy and quality	Malicious use	Security threats	Performance and 'explainability'	Strategic
Customer journeys *(eg, chatbots for customer services)*	✓		✓			✓	✓
Concision *(eg, generating content summaries)*	✓	✓				✓	
Coding *(eg, generating or debugging code)*		✓		✓	✓	✓	
Creative content *(eg, developing marketing content)*	✓	✓		✓		✓	

[1]Intellectual property.

McKinsey & Company

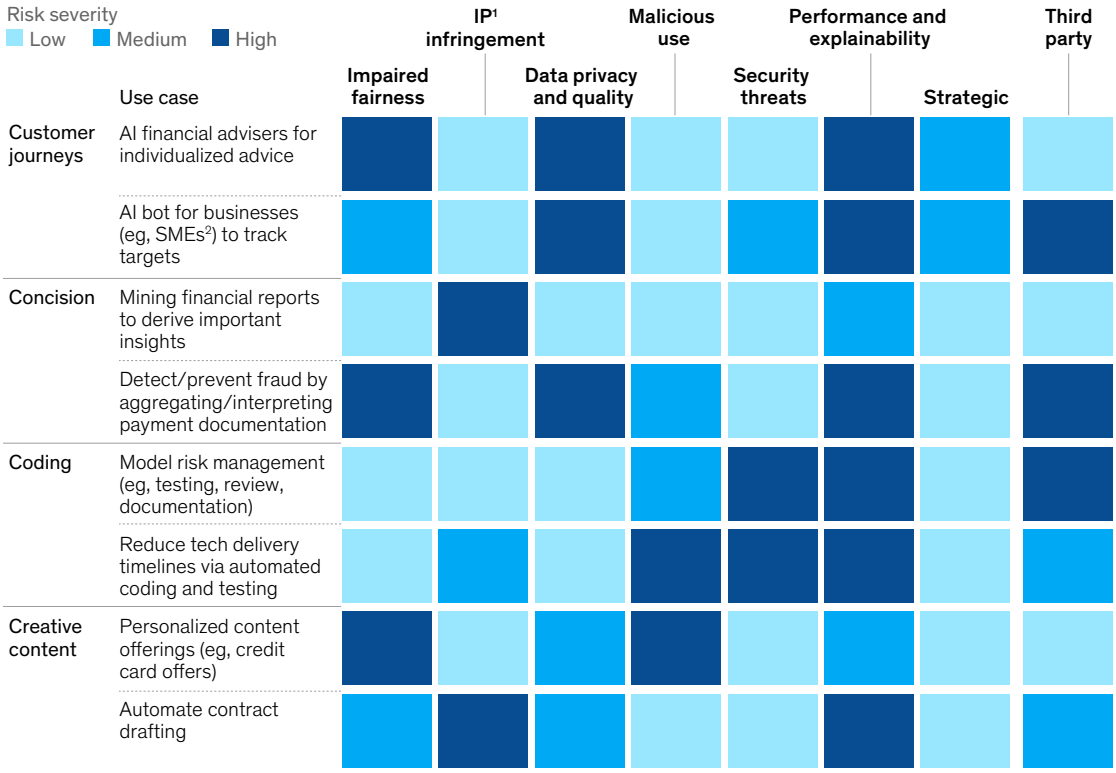
Identify risks across use cases

The essential starting point for organizations deploying gen AI use cases is to map the potential risks associated with each case across key risk categories to assess the potential risk severity. For example, use cases that support customer journeys, such as gen-AI-enabled chatbots for customer service, may raise risks such as bias and inequitable treatment across groups (for example, by gender and race), privacy concerns from users inputting sensitive information, and inaccuracy risks from model hallucination or outdated information (Exhibit 3).

When conducting this analysis, it's important to develop a rubric to calibrate expectations of what constitutes a high versus a medium risk across categories. Otherwise, organizations may run into disagreements driven more by individual comfort on risk levels than by objective factors. To take the example of data privacy, we typically see higher-risk examples as requiring personal or sensitive information for accurate training of the model (or higher potential for users to enter personal information in interacting with the technology). Lower-risk use cases would exhibit neither of these characteristics.

Exhibit 3

Organizations that deploy generative AI use cases can create a heat map ranking the potential severity of various categories of risk.



[1]Intellectual property.
[2]Small and medium-size enterprises.

McKinsey & Company

Using this logic, developing an application that supports an adviser in providing tailored financial advice would tend to rank higher in privacy risk exposure than would an application that automates basic contract templates.

It's essential that the executive in charge of the use case leads the initial assessment of the risks associated with it (as part of the role of the product manager in an effective operating model). This fosters the appropriate awareness of potential risks and accountability for managing them when the use case is approved for ultimate development. In addition, a cross-functional group, including business heads and members of legal and compliance functions, should review and validate the risk assessments for all use cases—and use the results as input when making decisions about use case prioritization.

Consider options for managing risks at each touchpoint

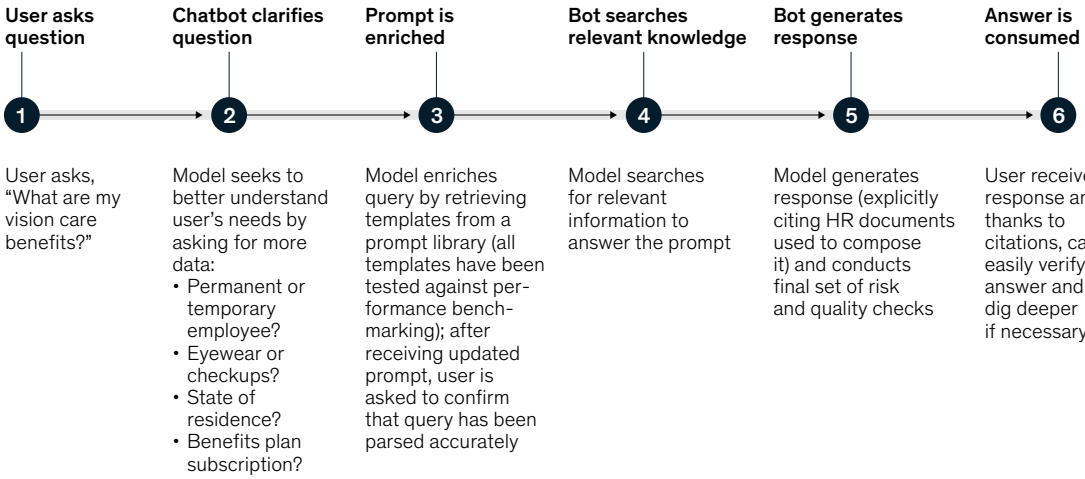
Once an organization maps the gen-AI-related risks, it must develop strategies to manage exposures through a combination of mitigation and robust governance. Many (but not all) mitigations are technical in nature and can be implemented across the life cycle of the process. Importantly, these controls don't all need to be embedded in the underlying foundation model itself (which many organizations won't have access to). Some can be overlays built in the local environment, as is the case of a gen-AI-enabled chatbot designed by an HR department to field employee queries about benefits (Exhibit 4).

In that use case, across the life cycle of a query, once a user asks a question, many possible mitigations can occur. They include having the chatbot ask clarifying questions to generate additional necessary user inputs, having the user confirm that the chatbot has properly understood the query, limiting the types of data sets that the chatbot can access (for example, excluding personal information), and designing the chatbot to provide citations to explain its answers and allow for fact-checking of its responses. Organizations implementing this use case can take steps (such as limiting repeated interactions) to frustrate the attack vectors and jailbreaking that are known to create challenges for chatbots. They can also develop classifiers to identify and reject out-of-scope queries (such as requesting calculations).

Exhibit 4

Generative AI risk can be mitigated at multiple points across a user interaction.

Sample HR chatbot interaction with built-in checkpoints to catch potential misfires



McKinsey & Company

There are important categories of additional non-technical mitigations that organizations should consider when developing use cases. At this stage of gen AI maturity, most organizations are maintaining humans in the loop to guard against the technology being able to put outputs directly into production or to engage directly with end customers. As previously referenced, contractual provisions to guard against problematic use of data from third parties are important. As a third example, organizations should develop coding standards and libraries to capture appropriate metadata and methodological standards to support reviews.

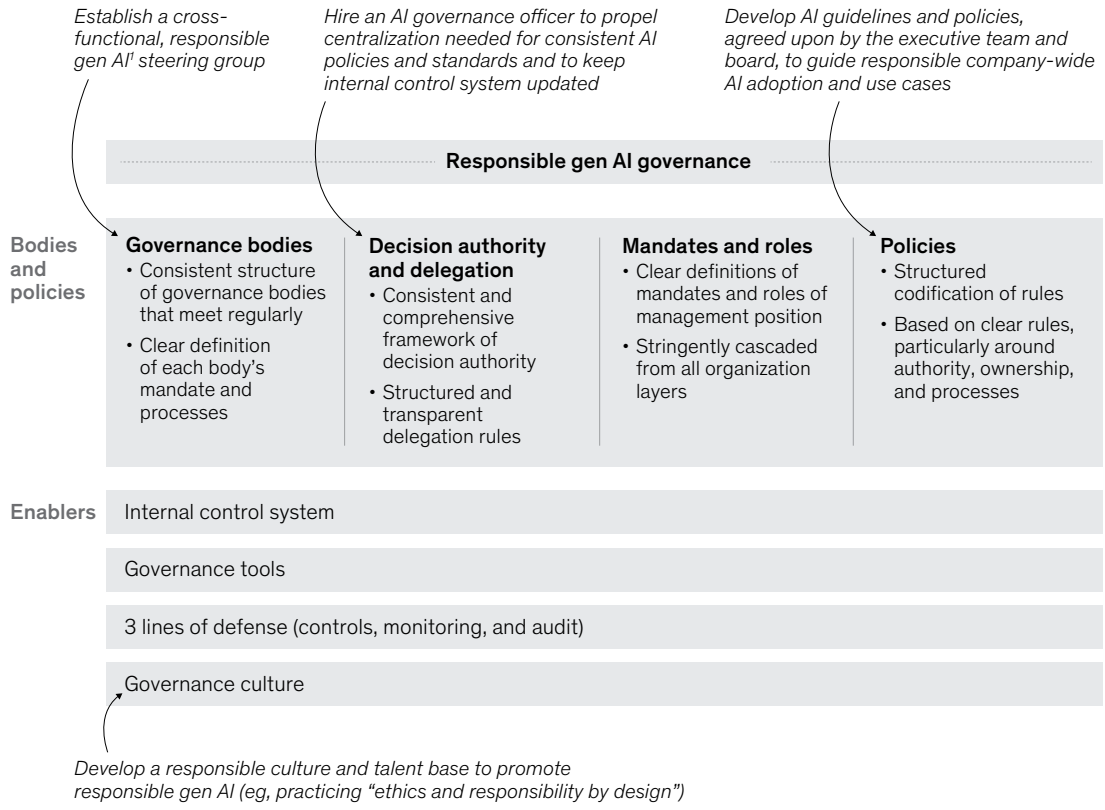
Many of the initial mitigating strategies for gen AI span multiple use cases, allowing organizations to get scaled benefits from their technical mitigations rather than having to create bespoke approaches for each case. For example, in the HR chatbot example, the ability to produce sources as part of the query answer could also be applied in use cases of an employee trying to explain a product to a customer or building analyses of peer companies. In both cases, this will go some way to addressing challenges of "explainability" and overall confidence in output.

Balancing speed to scale with judicious risk management through governance

Using gen AI will place new demands on most organizations to adapt governance structures to respond to demands on approvals and exercise oversight. However, most organizations should be able to adapt what they have today by expanding mandates or coverage (Exhibit 5). This will limit the

Exhibit 5

Moving with speed while mitigating risk often requires revised governance.



[1]Generative AI.

McKinsey & Company

potential disruption of establishing an entirely new phalanx of committees and approval bodies that could add friction to decision making and confusion over accountability.

Gen AI will likely require organizations to make changes to three core elements of governance:

— *A cross-functional, responsible gen AI steering group with at least a monthly cadence.* This group should include business and technology leaders, as well as data, privacy, legal, and compliance members. It should have a mandate for making critical decisions on managing gen AI risks, covering assessment of exposures and mitigating strategies for both inbound and adoption-based risks. It should review foundational strategy decisions, such as the selection of foundational models and compatibility with the organization's risk posture. This steering group ideally has a single individual empowered to handle coordination and agenda setting. In industries with established regulatory expectations and a long history of risk management of model and algorithmic risk (such as financial services), this person will typically be already on staff (and may be the head of model risk). For organizations facing a sudden increase in regulatory expectations from gen AI, they may need to hire an AI governance officer or similar role to discharge these responsibilities.

— *Responsible AI guidelines and policies.* Organizations should develop a set of guiding principles agreed on by the executive team and the board that will guide AI adoption and serve as a guardrail for acceptable use cases. Principles that we've seen debated include questions on the degree to which gen AI can or should be used to drive personalized marketing or customer outreach, the use of gen AI to support employment decisions (including hiring and performance reviews), and the conditions under which gen AI outputs can be put directly into production without human review. Existing policies typically need to be refreshed to account for gen AI development and use (for example, covering misrepresentation and IP infringement).

— *Responsible AI talent and culture.* A commitment to responsible AI can't rest solely in the executive ranks. Instead, it needs to cascade throughout the organization, with accountability, capability building, and awareness tailored to the relevant degree of exposure of relevant roles to the technologies. Basic organization-wide training on responsible AI should be developed and rolled out to foment a broad understanding of the dynamics of inbound risk and how to engage with the technology safely. For example, given the potential for the models to hallucinate, users should be told, as part of their training, that they shouldn't accept an answer just because their machine has provided it (in contrast to how they may have experienced prior office productivity technologies). Those engaged in the development and scaling of use cases should have a deep understanding of ethics and "responsibility by design" to embed risk considerations early in the design and engineering processes. Talent considerations include embedding a mix of nontechnical and technical talent—and ideally, technical talent with risk expertise to support identification and design of user query workflows and controls.

Implementing responsible gen AI: It's all about governance and people

Establishing the right governance is a necessary but not sufficient step in driving responsible adoption of gen AI use cases at scale. As referenced in the preceding section, embedding responsibility by design into the development process is essential for judicious deployment of the technology. There are four critical roles required for successful implementation of this throughout the use cases, where the responsibilities of these roles are tied closely to their talent and expected actions in pushing forward use cases:

— *Designers.* Designers, or product managers, steer the direction of gen AI deployment by identifying new use cases with an awareness of how they fit into the organization's overall gen AI strategy and road map. They're typically drawn from within the businesses and functions for which the organization has the most

conviction that gen AI can have significant impact. The product managers should be accountable for identifying and mitigating relevant risks. They will have an important role in driving the cultural changes required to adopt gen AI, including building trust in the proposition that business value can be achieved responsibly and safely for employees and customers.

— *Engineers.* Engineers are technical experts who understand the mechanics of gen AI. They develop or customize the technology to support the gen AI use cases. Just as important, they're responsible for guiding on the technical feasibility of mitigations and ultimately coding the mitigations to limit risk, as well as developing technical-monitoring strategies.

— *Governors.* Governors make up the teams that help establish the necessary governance, processes, and capabilities to drive responsible and safe implementation practices for gen AI. These include establishing the core risk frameworks, guardrails, and principles to guide the work of designers and engineers and challenging risk evaluation and mitigation effectiveness (especially for higher-risk use cases). The AI governance officer is a prime example of this persona, although the role will need to be complemented with others, given the range of potential risks. These roles will ideally cover data risk, data privacy, cybersecurity, regulatory compliance, and technology risk. Given the nascency of gen AI, governors will often need to coordinate with engineers to launch "red team" tests of emerging use cases built on gen AI models to identify and mitigate potential challenges.

— *Users.* Users represent the end users of new gen AI tools or use cases. They will need to be trained and acculturated to the dynamics and potential risks of the technology (including their role in responsible usage). They also play a critical role in helping identify risks from gen AI use cases, as they may experience problematic outputs in their interactions with the model.

An operating model should account for how the different personas will interact at different stages of the gen AI life cycle. There will be natural variations for each organization, depending on the specific capabilities embedded in each of the personas. For example, some organizations will have more technical capabilities in designers, meaning they may have a more active delivery role. But the intent of the operating model is to show how engagement varies at each stage of deployment.

Gen AI has the potential to redefine how people work and live. While the technology is fast developing, it comes with risks that range from concerns over the completeness of the training data to the potential of generating inaccurate or malicious outputs. Business leaders need to revise their technology playbooks and drive the integration of effective risk management from the start of their engagement with gen AI. This will allow for the application of this exciting new technology in a safe and responsible way, helping companies manage known risks (including inbound risks) while building the muscles to adapt to unanticipated risks as the capabilities and use cases of the technology expand. With major potential uplift in productivity at stake, working to scale gen AI sustainably and responsibly is essential in capturing its full benefits.

Oliver Bevan is a partner in McKinsey's Chicago office; **Michael Chui** is a partner in the Bay Area office, where **Brittany Presten** is an associate partner and **Lareina Yee** is a senior partner; and **Ida Kristensen** is a senior partner in the New York office.

This article was edited by Larry Kanter, a senior editor in the New York office.

Copyright © 2024 McKinsey & Company. All rights reserved.

McKinsey Risk & Resilience Practice

Global coleader and North America
Ida Kristensen
Ida_Kristensen@McKinsey.com

Global coleader and Europe
María del Mar Martínez
Maria_Martinez@McKinsey.com

Asia–Pacific
Akash Lal
Akash_Lal@McKinsey.com

Eastern Europe, Middle East, and North Africa
Luís Cunha
Luis_Cunha@McKinsey.com

Latin America
Elias Goraieb
Elias_Goraieb@McKinsey.com

Chair, Risk & Resilience Editorial Board
Thomas Poppensieker
Thomas_Poppensieker@McKinsey.com

Coleaders, Risk Knowledge
Luca Pancaldi and Sebastian Schneider
Luca_Pancaldi@McKinsey.com and
Sebastian_Schneider@McKinsey.com

In this issue

March 2024
Designed by LEFF
Copyright © McKinsey & Company
McKinsey.com