



How risk and compliance can accelerate generative AI adoption

Businesses are keen to capture the benefits of generative AI. Too often, however, risk concerns are stalling progress. A robust AI risk review process can help.

Introduction

It has been almost a year since generative AI exploded our understanding of what artificial intelligence could do for business. The emerging technology's ability to consume and organize vast amounts of information, mimic human understanding, and generate content quickly created enormous expectations for a surge of technology-led productivity growth. In a KPMG survey of top executives, more than 70 percent of respondents said they expected to implement a generative AI solution by spring 2024 and more than 80 percent expected the technology to have "significant impact" on their businesses by the end of 2024.

But in many organizations, generative AI plans are stuck or progressing slowly as businesses struggle to account for all the risks—security, privacy, reliability, ethical, regulatory, intellectual property, etc. This is where the risk function—risk, compliance, and legal teams—can step in and play a critical role. By developing and activating a process to quickly assess and control risks around generative AI models and data sets, risk teams will become an enabler for the business rather than a speed bump that limits agility.

In this paper, we will look at how risk functions can help their organizations move forward in generative AI adoption by defining the scope and severity of the risks; aligning principles, processes, and people to manage them; and establishing a durable, practical framework for trusted AI governance. Our focus here is on generative AI, but Trusted AI and other governance practices that we discuss here also apply to how risk can support the organization's wider AI agenda.

A rising risk awareness

Generative AI presents an array of multilayered and multidisciplinary risk issues—in some cases, introducing pure black-box unknowns. These challenges will require new depth, expertise, and leadership from the risk function. Corporate leaders are well aware of generative AI’s operational, regulatory, and reputational risks, and they are looking to develop new approaches and solutions.

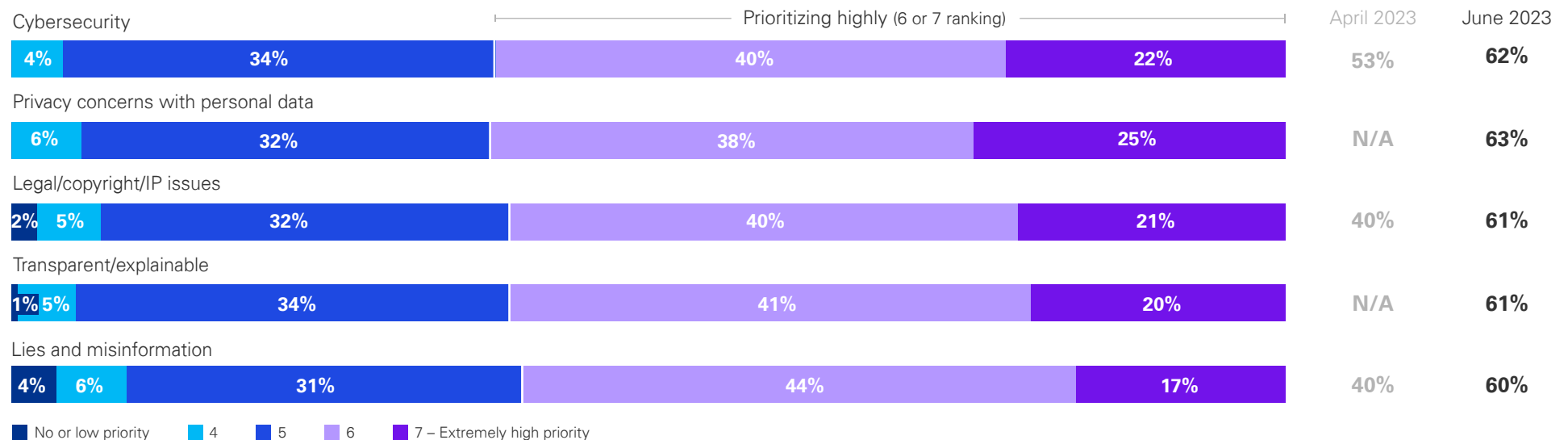
What risk concerns matter most? In an April 2023 KPMG survey of top leaders, 40 percent of respondents said legal and copyright issues were a chief risk concern. In a second KPMG survey in June 2023, 61 percent said these were top concerns, a greater than 50 percent increase. There were similar increases

in concerns over misinformation, inaccuracies, and weaponization of generative AI content (Exhibit 1).

As organizations begin to adopt generative AI, there is a strategic imperative to systemize risk management to enable ethical and responsible deployment of AI at scale, respond efficiently to changes in the regulatory environment, and maintain stakeholder trust. To do this, it is critical to integrate risk management into the design of AI models—introducing the risk and compliance mindset at the start of the model lifecycle and maintaining a focus on risk through implementation, optimization, and use.

Exhibit 1. Leaders have a growing awareness of generative AI risks

Prioritization of risk management



Source: KPMG executive surveys. Note: Totals may not sum due to rounding.

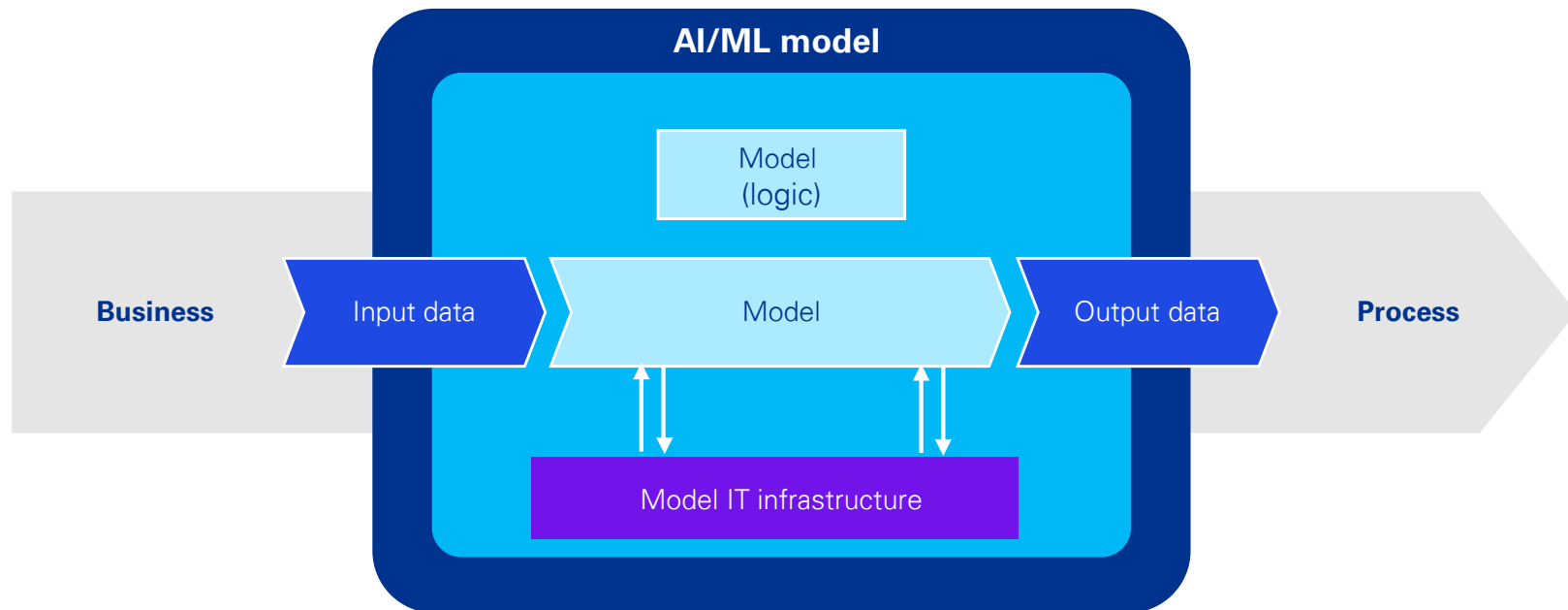
Q19. What level of priority is your organization placing on risk management and mitigation in the following areas to maintain the trust of your stakeholders when it comes to Generative AI? Please rank each on a scale of 1-7.

Getting started: Classifying generative AI risks

While the level of risk awareness has risen, leaders also believe their risk professionals are prepared to deal with the risk challenges posed by generative AI. More than three-quarters of respondents in the June survey said they were highly confident in their organization's ability to address and mitigate risks associated with generative AI. However, this is not the reality we see on the ground. We find that many organizations remain stuck, searching for leadership, consensus, and a rational approach for resolving generative AI risk issues and establishing guardrails for ongoing protection. The risks associated with generative AI fall into four buckets:

- 1 **Model context and governance risks**
- 2 **Input data risks**
- 3 **Output risks**
- 4 **Model logic and infrastructure risks**

Exhibit 2. How an AI model fits into a business process



Model context and governance

Understanding AI model context and governance is paramount to any generative AI deployment. The context defines the purpose, scope, and ethical framework within which the generative AI model operates. It also defines the required training data, its sources, and model architecture and determines applicable risks (security, fairness, sustainability, etc.). A model using sensitive information or information used for financial reporting, for example, will have additional organizational and regulatory requirements that need to be addressed.

Governance provides the structure to manage and oversee the model's development and usage. Failure to understand the model's context and establish robust governance can result in ethical lapses, regulatory violations, unauthorized access to sensitive or confidential data, unintended biases, and misalignment with organizational values. Effective governance outlines the core principles and expectations to be addressed and allows flexibility and adaptability, depending on the risk associated with a given AI model or use case.

Input data

The quality, relevance, and fairness of input data directly impact the model's effectiveness and ethical hygiene. Using high-quality and diverse data sets in training and fine-tuning generative AI models also can help reduce or prevent hallucinations. In accordance with the defined context, organizations need to ensure that the data sources and pipelines used by generative AI models for training, validation, and inference are

free from biases and that they are trustworthy. Applicable principles such as data integrity, security, privacy, and transparency should be considered to ensure and protect the integrity of the data, provide clarity of data usage, and obtain appropriate consent. To help enforce these principles, for example, organizations can implement continuous monitoring of data sources to help ensure input data issues are identified.

Output data

Generative AI can produce dazzling results—well-argued legal briefs, insightful reports, and in-depth analyses. However, before organizations trust these generative AI outputs, they need to ensure that they have robust quality assurance practices—both manual and automated—in place to check and

fix results. The quality and completeness of input data, the model logic, and infrastructure effect the outputs of a generative AI model.

Without the right measures in place, outputs could be biased, unethical, irrelevant, or incoherent, or expose sensitive data, leading to legal and safety

consequences. Mechanisms such as content filtering and moderation, alerts and anomaly detection, and drift detection should be considered. Regular data quality and validation audits, ethical

and fairness audits, and model retraining and fine-tuning should be performed to verify that output data continues to meet organizational goals and values.

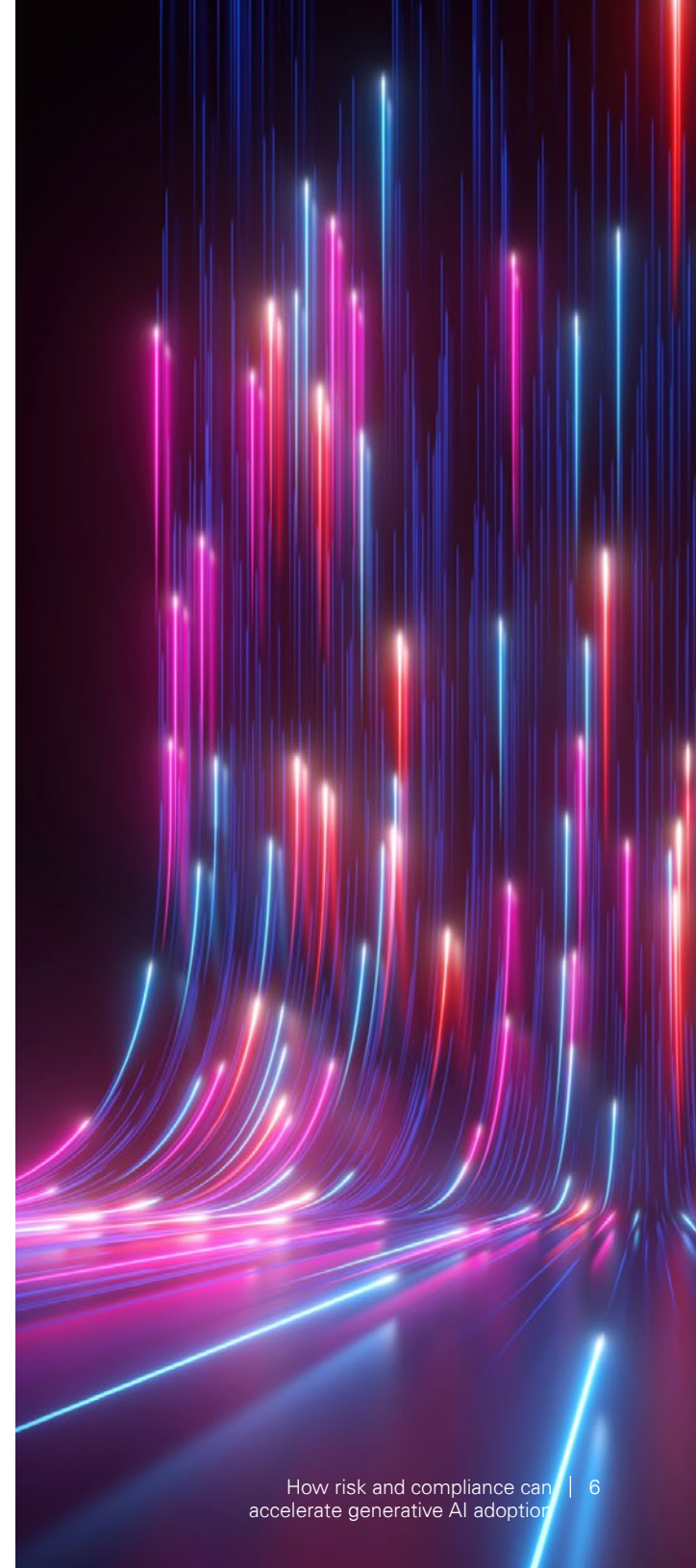
Model logic and infrastructure

The model logic defines how the model operates, processes data, and generates responses or outputs. Will the AI application only look for absolute matches in the data provided in generating responses? Or will the AI application be able to “guess” the response? Or will the model access information from the internet? Answers to these questions will be driven by the context of the application and setup of the model logic leading to a direct influence on the outcomes and decisions of a generative AI model. An understanding of the data flow is critical to evaluating the model ecosystem and ensuring successful outcomes.

The infrastructure provides the necessary computational resources and environment for execution, underpinning the model’s functionality and scalability. Understanding and selecting the right infrastructure can help organizations choose cost-effective hardware and software resources while meeting performance, reliability, flexibility, and scalability goals. Careful infrastructure design also enables organizations to future-proof their environment.

The generative AI technology stack is built on familiar foundations, including the application layer, AI models, and hosted cloud infrastructure. Understanding the Shared Responsibility Model (SRM) and the specific risk management implications for the service provider and the customer is critical. An incomplete understanding of SRM can lead to performance bottlenecks, security vulnerabilities, and operational challenges. Organizations should consider establishing stringent controls, including model testing, vulnerability assessments, scalability planning, and incident response protocols.

To monitor infrastructure on an ongoing basis, organizations should consider other security and service management processes, such as access management, API (application program interface) management, configuration management, and change management. These controls, on top of a clear understanding of the model logic and infrastructure, will help ensure that the generative AI model functions reliably, securely, and at scale.



How risk can enable generative AI adoption

Business leaders are looking to generative AI to quickly increase efficiency and enable new sources of growth. In a KPMG survey, top leaders cited competitive pressures as a chief reason to adopt generative AI in operations. They also expect generative AI to open new possibilities in sales, marketing, and product development. And they are eager to get going so they do not fall behind.

Whether the generative AI solution is developed in-house, provided by a third party, or is a hybrid model, involving risk functions at the outset helps expedite the approval process.

Working with other functions—information technology (IT), cyber, finance, human resources (HR), etc.—risk can help shorten the path to generative AI implementation. We have identified four key risk areas that need to be addressed

before any implementation proceeds (Exhibit 3). These are just some of the elements of the larger AI risk program that will provide ongoing oversight.

To expedite reviews and approvals on these risk topics, companies can set up a multidisciplinary generative AI task force. When a business unit or function proposes a generative AI implementation, finance can assess its potential payoff and strategic priority while IT will need to determine how it can be implemented—including choosing vendors/partner and supporting large language model (LLM).¹ Then risk should swing into action, using a streamlined, repeatable governance and risk review process that should include collaboration with legal, compliance, and cybersecurity departments.²





Another step risk can take early in the adoption process is establishing a “risk sanitized” version of a base LLM that the businesses can use without going through a broad review. By identifying and clearing acceptable levels of risk on this model, the risk function can expedite the project approval process by focusing only on the particular use case and its specific customizations rather than on the approved broader model.



¹ See “[Why finance should lead the adoption of generative AI](#)” and “[The CIO’s path to driving value with generative AI](#),” KPMG LLP, 2023

² See “[Using generative AI to strengthen cybersecurity](#),” KPMG LLP, 2023

Exhibit 3. Critical risk issues for generative AI projects

	Strategy and design	Data	Modeling	Evaluation	Deployment and optimization
 <p>Sustainability</p>	<p>Risk of insufficient sustainability considerations</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Define minimum sustainability requirements Environmental impact indicators 	<p>Risk of energy-intensive data handling</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Minimize data collection 	<p>Risk of model sustainability degradation</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Choose efficient algorithms Choose energy-optimized hardware 	<p>Risk of sustainability practice implementation errors</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Evaluate model environmental impact 	<p>Risk of not adhering to organizational environmental, social, and governance (ESG) commitments</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Monitor energy consumption
 <p>Privacy</p>	<p>Risk of privacy considerations not being identified prior to development</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Incorporate privacy by design Develop privacy metrics 	<p>Risk of violating data holders' privacy rights at collection</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Adhere to privacy practices at collection Mask and/or minimize data at collection 	<p>Risk of deviating from the organization's privacy principles and commitments</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Develop models that enable and adhere to privacy principles 	<p>Risk of privacy practice implementation errors</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Evaluate against privacy metrics 	<p>Risk of violating international privacy regulations</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Monitor regulatory changes Monitor against privacy metrics
 <p>Transparency</p>	<p>Risk of misrepresenting model features and/or use to stakeholder groups</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Identify, involve, and document transparency goals for relevant stakeholders 	<p>Risk of violating consumer trust and/or rights in data collection</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Inform impacted individuals of data collection 	<p>Risk of perpetuating nontransparent use practices</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Develop mechanism(s) to inform users of interaction and/or use of generative AI 	<p>Risk of transparency practice implementation errors</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Evaluate model against transparency goals 	<p>Risk of losing stakeholder trust through nontransparent use practices</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Monitor model against transparency goals Publish transparency to stakeholder groups
 <p>Fairness</p>	<p>Risk of designing unethical and/or biased models</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Identify and consult diverse groups of internal and external stakeholders on strategy 	<p>Risk of collecting nonrepresentative data</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Define and follow guidelines for fair data collection practices 	<p>Risk of biases being embedded into model features</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Consult diverse stakeholders to identify potential sources of bias 	<p>Risk of not identifying biased outcomes</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Evaluate performance against fairness metrics Diverse groups 	<p>Risk of perpetuating and amplifying social biases</p> <p><i>Control examples:</i></p> <ul style="list-style-type: none"> Obtain model feedback from diverse stakeholder groups

Build a Trusted AI framework

Organizations that are planning to use generative AI in their business must understand the importance and strategic imperative of ensuring that all AI applications, including those using generative AI, are trustworthy and responsible. Reputations are at stake, and without governance in place to ensure the technology is operating ethically and reliably, businesses risk damaging their relationships with customers, employees, partners, and the market.

The AI risk landscape calls for an agile, rapid, granular, and focused response grounded in a collective thought process from

stakeholders across the organization. This is where an organization's risk functions can step in. Risk functions should be empowered to help ensure that AI development and deployment aligns with ethical and legal principles while being accountable and transparent to stakeholders. This is where creating and operationalizing your Trusted AI framework comes in. KPMG has developed a Trusted AI framework that stresses fairness, transparency, explainability, accountability, data integrity, reliability, security, safety, privacy, and sustainability (Exhibit 4).

Exhibit 4. The Trusted AI framework

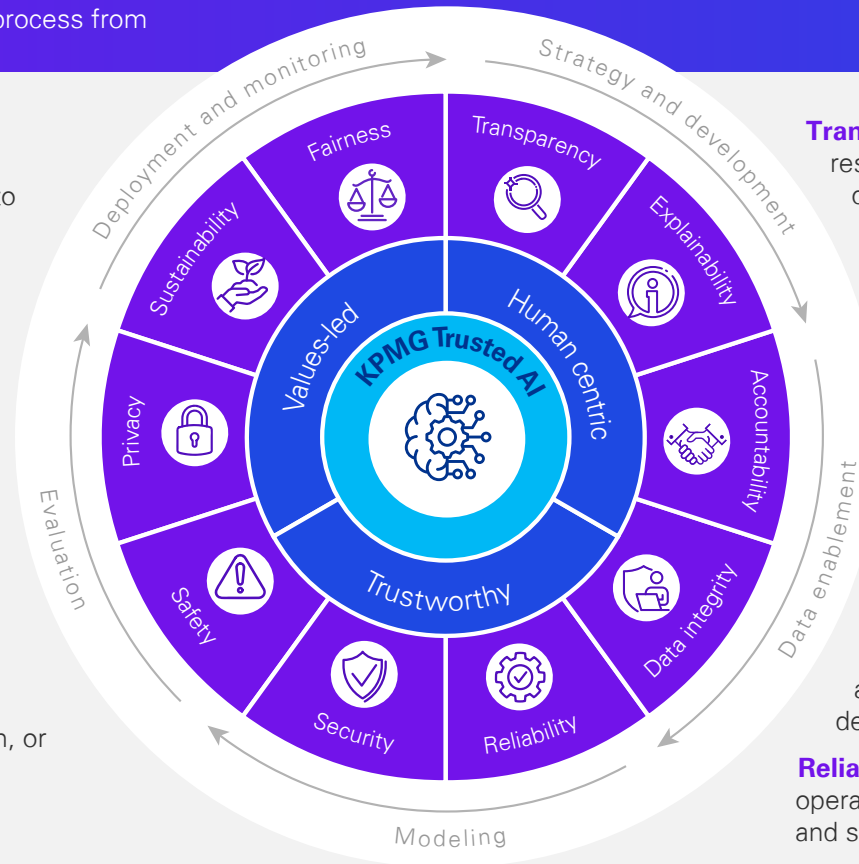
Fairness. AI solutions should be designed to reduce or eliminate bias against individuals, communities, or groups.

Sustainability. AI solutions should be designed to be energy efficient, reduce carbon emissions, and support a cleaner environment.

Privacy. AI solutions should be designed to comply with applicable privacy and data protection laws and regulations.

Safety. AI solutions should be designed and implemented to safeguard against harm to humans and/or property.

Security. Robust and resilient practices should be implemented to safeguard AI solutions against bad actors, misinformation, or adverse events.



Transparency. AI solutions should include responsible disclosure to provide stakeholders a clear understanding as to what is happening in the solution across the AI lifecycle.

Explainability. AI solutions should be developed and delivered in a way that answers the questions of how and why a conclusion was drawn from the solution.

Accountability. Human oversight and responsibility should be embedded across the AI lifecycle to manage risk and comply with applicable laws and regulations.

Data integrity. Data used in AI solutions should be acquired in compliance with applicable laws and regulations and assessed for accuracy, completeness, appropriateness, and quality to drive trusted decisions.

Reliability. AI solutions should consistently operate in accordance with their intended purpose and scope and at the desired level of precision.

The Trusted AI framework can be used across AI activities—establishing safe and ethical practices for machine learning and LLM teams and defining data quality standards. The Trusted AI framework provides an ethical and responsible “north star” for organizations as they build and expand their generative AI programs, helping organizations to manage the risks associated with AI without hindering or blocking deployments. By crafting processes and policies that bring Trusted AI to life throughout the entire deployment lifecycle, these principles become a part of the organizational DNA.

The central challenge in AI risk management lies in the dynamic nature of each AI application, which comes with its own business processes, impact, and risks. However, a Trusted AI framework can be a solid foundation that is adaptable to any generative AI application. Defining the framework is only the start; risk functions should also ensure that the framework is operationalized and adopted across the organization. Adoption and operationalization hinges on a framework that is not too theoretical and does not overly burden the business. KPMG, for example, uses commercially available software products to automate aspects of AI risk management and governance for itself and its clients.

Conclusion

In conclusion, identifying and managing risks early can only help companies implement AI solutions quickly, safely, and responsibly. Risk mitigation should be part of the full project lifecycle, which will enable mitigation solutions to be built into deployment plans up front. By treating risk as an afterthought, organizations end up retrofitting incomplete solutions or addressing costly rework to properly address issues such as storage of confidential information or privacy matters.

How KPMG can help

At KPMG, we bring extensive industry experience, the latest technical skills, and innovative solutions to every generative AI project. With our strong partner network, we empower business leaders to leverage the full potential of generative AI in a secure and reliable way. From developing initial strategies and designs to managing ongoing operations, we support our clients throughout the entire AI lifecycle. Our risk-management services

include rapid assessments of existing generative AI frameworks, benchmarking analysis, and implementing a robust governance process from intake to production.

Our Trusted AI framework helps ensure that AI implementation and usage are ethical, trustworthy, and responsible based on these principles:

1 Values driven



We will implement AI guided by our values. They are our differentiator at KPMG and shape an open and inclusive culture that operates to the highest ethical standards. Our values inform our day-to-day behaviors and help us navigate emerging opportunities and challenges. We take a purpose-led approach that empowers positive change for our clients, our people, and our communities.

2 Human centric



We will prioritize human impact as we deploy AI and recognize the needs of our people and clients. We are embracing AI to empower and augment human capabilities—to unleash creativity and improve productivity in a way that allows people to reimagine how they spend their days.

3 Trustworthy

We will adhere to our framework across the AI lifecycle and its 10 pillars that guide how and why we use AI. We will strive to ensure our data acquisition, governance, and usage practices uphold ethical standards and comply with applicable privacy and data protection regulations and confidentiality arrangements.

Related thought leadership



Visit our [Digital Trust webpage](#) for a library of the latest thought leadership.

Your contacts

Dr. Matthias Bossardt

Partner,
Head of Cyber & Digital Risk Consulting

E: mbossardt@kpmg.com
T: +41 58 249 36 98

Dr. Thomas Bolliger

Partner,
Cyber & Digital Risk Consulting

E: tbolliger@kpmg.com
T: +41 58 249 28 13

Michele Daryanani

Director,
Cyber & Digital Risk Consulting

E: micheledaryanani@kpmg.com
T: +41 58 249 61 74

KPMG AG

Badenerstrasse 172
PO Box
8036 Zürich

kpmg.ch/digitaltrust

Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates or related entities.



The information contained herein is of a general nature and is not intended to address the circumstances of any particular individual or entity. Although we endeavor to provide accurate and timely information, there can be no guarantee that such information is accurate as of the date it is received or that it will continue to be accurate in the future. No one should act on such information without appropriate professional advice after a thorough examination of the particular situation. The scope of any potential collaboration with audit clients is defined by regulatory requirements governing auditor independence. If you would like to know more about how KPMG AG processes personal data, please read our Privacy Policy, which you can find on our homepage at www.kpmg.ch.

© 2023 KPMG AG, a Swiss corporation, is a subsidiary of KPMG Holding AG, which is a member firm of the KPMG global organization of independent member firms affiliated with KPMG International Limited, a private English company limited by guarantee. All rights reserved.