KPMG

Deploying trustworthy AI
An illustrative risk and controls guide

The guide to AI risks and underlying control considerations for risk, technology, compliance, and legal leaders

visit.kpmg.us/TrustedAI

KPMG. Make the Difference.

Foreword

AI is on the rise. Controls can help manage the risks.

Artificial intelligence (AI) is revolutionising sectors, transforming business structures, and even altering our way of life and work. It also holds the potential to significantly reshape the future of your organisation.

The accomplishments enterprises can achieve with AI are seemingly limitless. According to the KPMG 2024 CEO Outlook, 64 percent of global CEOs say AI is a top investment priority, despite uncertain economic conditions with top expected benefits being increased efficiency and productivity, an upskilled workforce, and increased enterprise innovation.[1]

Unsurprisingly, such benefits make executives eager to integrate AI into their businesses and accelerate the value it delivers. **But organisations can only harness AI's full potential once they ground such initiatives in trust, managing its complexities and risks in a responsible, ethical, and transparent manner.** As the scale and complexity of AI adoption advances across business operations, such complexities become increasingly difficult to navigate.

The stakes are also rising for those tasked with ensuring the safe deployment and use of AI applications – risk and compliance departments, cyber and information security teams, data and privacy offices, legal teams, and internal audit. AI systems that are not properly governed and controlled can hinder returns on AI investments, lead to regulatory compliance violations, result in data and IP loss, or damage the organisation's reputation.

Ultimately, it will be key to ground AI systems in pragmatic and scalable risk management practices to **deploy AI boldly, quickly, and responsibly – unlocking its transformative benefits.** Establishing a robust risk and controls guide for managing AI risks is a critical step in developing an AI risk management program.

KPMG has published a first–of–its kind illustrative AI risk and controls consideration guide.

The guide – aligned to the KPMG Trusted AI framework – provides a structured approach for organisations to begin identifying AI risks and designing proportionate control considerations to mitigate those risks. While existing AI frameworks and standards identify risks at different stages of the AI lifecycle, this guide delves into the underlying control activities, outlining suggestive control considerations businesses should contemplate for managing AI risks.

Please note: This guide is meant to be an informative aid for helping organisations like yours appropriately manage AI–specific risks. It provides illustrative examples of potential control considerations to address a large, though not complete, set of AI–specific risks.

Intentionally focused solely on AI risks, it is designed to complement existing risk management frameworks that address general technology risks across domains such as security, data privacy, and third–party risk management. As such, you should first identify control considerations from this guide that are relevant to your business, and then carefully integrate them with your existing risk and control frameworks to help ensure a thorough view of risks across your organisation.

We hope that this guide helps your organisation begin to navigate the complex landscape of AI risks and drive innovation in a trusted manner.

[1] KPMG 2024 US CEO Outlook

How to put this guide into practice

Who is this guide for?

This guide can serve as a resource for any anyone leading or involved in AI risk management and governance, including risk and compliance departments, cyber and information security teams, data and privacy offices, legal teams, and internal audit.

Start with these questions.

How does the risk and related set of control considerations align to existing risk taxonomies in my business?

This guide is aligned to the 10 pillars of the KPMG Trusted AI framework, and was developed around leading AI frameworks and regulations, such as ISO 42001, the National Institute of Standards and Technology (NIST) AI risk management framework, and the EU AI Act. This is meant to be complementary to existing risk taxonomies within your organisation, such as IT general controls and data governance controls.

How should the control considerations be applied across the AI lifecycle?

To identify and implement control considerations across the AI lifecycle, there are several factors organisations should consider, such as understanding the nature and use of the AI system; data flow, configuration, and logic that influences operation; and learning types and data sources used.

How can we design and implement the control considerations to fit our own organisation and AI system?

Not every organisation or AI system may need to implement every control or there may be additional controls based on your specific deployments. Users of this guide should consider existing risk and control taxonomies in place and relevant to AI, such as IT general controls, data governance controls, access and security controls, application programming interface (API) controls, etc. Additionally, users should consider, for example, the nature of the AI deployments, and whether AI systems are third party, internally developed, leverage proprietary data sources, or have other configuration or techniques in play (such as retrieval augmented generation) which may influence risks and AI system operation. These considerations help to inform what risks may be present and, therefore, control activities required.

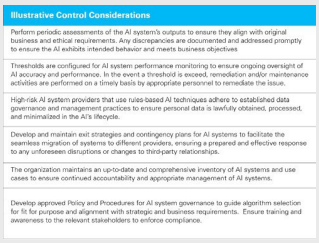
Get started by exploring the KPMG Trusted AI framework

01

Explore Trusted AI pillars

02

Determine relevant risk categories



03

Identify relevant control considerations

04

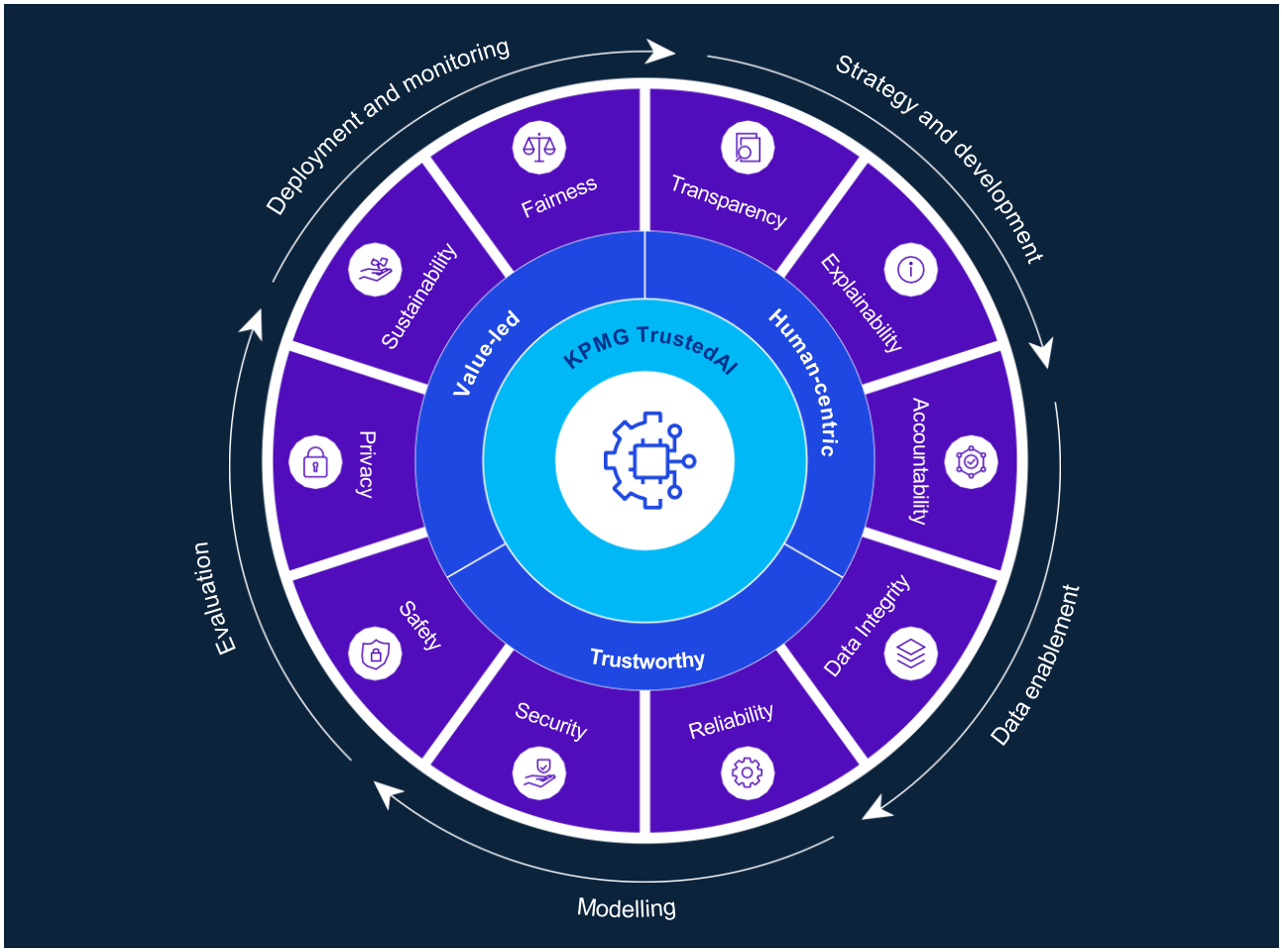
Determine relevant risk categories

Trusted AI pillars of risk and controls guide

About the KPMG Trusted AI framework

The AI Risk and Controls Guide is aligned to our Trusted AI framework, which is rooted in a values–driven, human–centric, and trustworthy approach to AI development and deployment. The Trusted AI framework helps our own firm, and our clients, develop and deploy AI solutions that address ethical concerns and comply with regulatory standards.

Organised under the 10 pillars of the KPMG Trusted AI framework, this guide outlines an initial inventory of AI risks, each with a set of control considerations that organisations can leverage as they build out their control catalogues.



Transparency

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

Risk Categories

Distinguishing Human vs. AI Content

Failure to distinguish between human–generated and AI – generated content can lead to misinformation, confusion, compromise the integrity of information sources, and/or lead to consumer mis – trust.

Risk Categories

Lack of Transparency in AI and Data Usage

Lack of transparency in AI and data usage can undermine user privacy, cause unaccountability for errors or harm, and the potential to violate ethical standards, thereby eroding public trust in such technologies.

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.

Risk Categories

Explainability Not Embedded in the Design

AI systems are not designed, developed, or implemented with explainability principles in mind—when explainability is not considered at the start of the AI lifecycle, the result is solutions with profound downstream implications on system use, trust, and performance.

Lack of Meaningful Human Review or Intervention

Humans need to be aware of the use of AI, provide oversight, and be able to override decisions made by AI systems.

Accountability

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

Risk Categories

AI Performance Erodes Over Time

Inability to identify and monitor the use of AI systems' performance may result in the erosion of performance over time.

Bypassing AI Risk Management

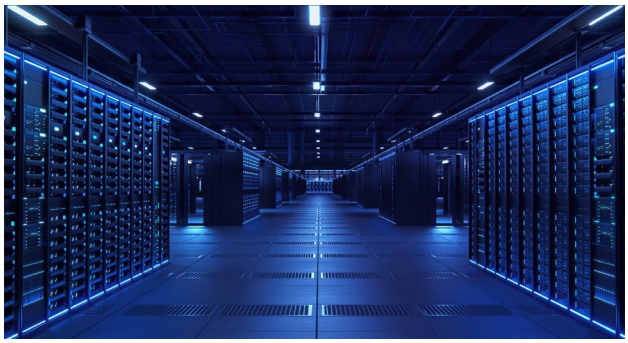
Development and use of AI tools without proper oversight can expose the enterprise to risk.

Ineffective AI Lifecycle

Lack of ownership of AI tools throughout the lifecycle can cause AI to drift from organisational strategy and intended objectives.

Organisational Accountability

A lack of accountability over AI systems may result in noncompliance with organisational and/or regulatory requirements.



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.

Risk Categories

Lack of Data Integrity in AI Systems

Compromised data integrity in AI systems may lead to inaccurate or unreliable outputs, undermining decision – making processes and potentially causing operational and reputational harm.

Reliability

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

Risk Categories

Insufficient Support and Maintenance

Insufficient operational support and maintenance leads to an ineffective AI solution, or to AI solutions becoming ineffective over time, and/or poor decision – making during major incidents.

Insufficient Understanding of AI Architecture

IT and data components of the overall AI environment, including changes to IT infrastructure, AI models, algorithms, and data, may not be fully understood.

by the operational IT support at the organisation, undermining the reliability and robustness of the AI systems and potentially disrupting the continuity and smooth operation of the overall business.

Security

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

Risk Categories

AI Security

Failure to embed security principles in the AI model architecture and AI development processes can lead to significant security vulnerabilities and/or unauthorised disclosure of information (including Personal Data and Intellectual Property).

Unsafe Prompt Engineering

Prompt engineering may result in unintended consequences including, but not limited to, leaks of strictly confidential information/Personal Data, creation of malicious code, social engineering, or system outages.

Safety

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

Risk Categories

Inadequate Response to AI – Generated Safety Threats

Organisational procedures and systems are insufficiently robust to quickly and effectively respond to safety threats generated or exacerbated by AI systems, leading to potential harm or hazardous situations.

Threat to Humans

AI systems may be leveraged or misused as a threat to human life and well – being, resulting in potential harm or adverse effects on society.

Privacy

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

Risk Categories

Privacy Violations from AI Solutions

Failure to comply with Organisation Privacy Directives and Procedures (e.g., inappropriate collection/disclosure of personal data) may result in a loss of consumer trust, regulatory non – compliance, or cause financial harm.

Sustainability

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

Risk Categories

Overarching Risk Associated with AI Sustainability

Lack of a sustainable AI strategy, efficient energy consumption, and understanding of e – waste generation may result in negative environmental, ethical, societal, and operational impacts.

Fairness

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

Risk Categories

Harmful Bias in AI Systems

Harmful bias in AI systems can perpetuate societal inequalities or discriminatory outcomes, which may lead to the erosion of public trust and cause legal, reputational, or financial loss.

How KPMG can help

The KPMG Trusted AI framework offers a pathway to help harness AI's potential in a trusted manner, and our suite of AI Trust services and solutions helps companies put the framework into action.

Our services include:

Trusted AI strategy

Assist organisations in assessing their current AI capabilities and crafting strategic roadmaps that enhance potential.

AI ethics and governance

Assist in the development of robust AI governance frameworks, controls, and operating models to help ensure AI is trustworthy. This includes comprehensive risk, policy, and controls assessments, alongside AI regulatory compliance.

AI risk assessment and regulatory compliance

Help organisations assess where they are in their Trusted AI journey by conducting risk – based AI assessments across AI use cases. This includes AI readiness, maturity assessments, AI strategy review, and assessing consistency of AI solutions with evolving frameworks and regulations.

Machine learning operations

Develop leading constructs, processes, and technologies for model management to help build trust in AI models, supporting their governance, lifecycle management, and effective deployment and monitoring.

AI security

Provide strategies, processes, and tools to help enhance AI security and privacy, helping organisations detect, respond to, and recover from cyber threats, privacy risks, and adversarial attacks.

AI assurance

Help test, examine, and report on the management processes, controls, and claims regarding the responsible use of AI technologies:

- AI assurance scoping
- AI diagnostics reviews
- AI model control testing

For more information:
visit.kpmg.us/TrustedAIservices

Need a customised AI Risk and Controls Guide?

KPMG can help customise and tailor the AI Risk and Controls Guide to meet the specific needs and challenges of your organisation, provide targeted training and education to help ensure a deep understanding and effective application of the matrix's principles, and deliver ongoing support and advisory services to navigate emerging AI risks and opportunities. Specific services we offer that can help your team tangibly implement the framework include:

- AI governance design and operations support: establishing or enhancing your AI governance program, policy, and operating model, or helping to scale and operationalise your AI governance program.
- Regulatory mapping: mapping to existing taxonomies to help ensure a complete control portfolio.
- Lifecycle mapping: aligning controls that best fit to different stages of the AI lifecycle.
- Control implementation support: documentation, design, and implementation support for AI controls.
- AI assessments: conducting AI assessments, compliance assessments, or risk – based governance assessments.

kpmg.com/uk

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

CREATE: CRT166904A | March 2026