# Responsible & Trusted AI

Richard Knight & Chris Jambor

# What is impacting the market?

# Common challenges the C-suite at our clients are facing

## How do organizations safely and responsibly unlock value from AI - and achieve their business ambitions?


**Secure our models** from adversarial attacks

**Ensure compliance** with global AI regulations

**Harness the value** of our AI at scale and responsibly

**Protect ourselves** from financial and reputational risks

**Enhance the trust of our consumers** (internal, external)

**Drive accountability and transparency**

# Scaling AI has introduced a growing number of challenges

## Security and privacy

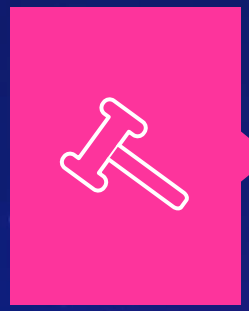
The use of Generative AI poses security and privacy risks, which could result in data breaches, reputational damage, or privacy regulation violations. Increasing sophistication from threat actors and velocity of malware and cyber attacks.

## Data quality, integrity, and bias

Generative AI presents potential risks to data quality, integrity, and bias. If not managed properly, it could result in inaccurate or biased outcomes, leading to legal liabilities, loss of client trust and reputational damage.

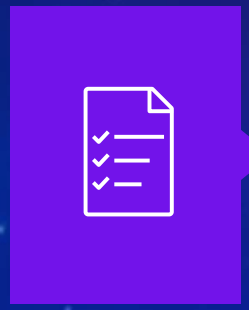
## Intellectual property

Lack of legislation defining ownership of AI generated content may result in the inability to obtain copyright of content produced. Additionally, unclear terms of use may result in unintended violation of intellectual property rules.



## Regulatory and professional standards

Regulators have not provided clear guidance on the use of Generative AI. Navigating regulatory requirements and adhering to our professional standards may pose challenges due to unclear guidance.



## Policy

Organizations must amend existing IT policies by identifying scenarios for use, aligning with data governance and ethical standards, and provide adequate training to users. Failure to do so may result in policy violations, legal liabilities, and ethical concerns.

## Brand and marketing

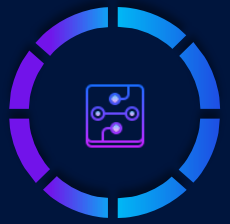
May perpetuate or amplify existing biases in the marketing and branding. Can result in negative impact on brand image and market share. An overreliance on AI generated content may lead to a lack of creativity and originality in marketing campaigns.

# Evolving AI Technology Trends

## Multi-Modal LLMs

LLMs that can understand and generate different types of data, such as text, images, video, code, and audio. E.g: hand sketch to code.

*Unleashed Power of Multi-Sensory AI*



## Embedded AI

AI seamlessly integrated into SaaS and other apps & platforms. Seamless and automation / augmentation and intelligence. E.g.: Einstein GPT

*More efficiency, automation & transformation*

## AI Agents

Programs that operate autonomously in an environment. Proactive. Have memory. Take actions to achieve specific goals. E.g.: Meeting Assistant

*Copilots for Humans. Humans as Copilots*

## Expanding Ecosystem

More democratization of AI. Growing set of models, tools, libraries, & talent open up AI development to everyone, not just experts.

*Democratized access to AI*

## Trustworthy AI

Regulations and evaluation criteria coming. Integrated Risk, Compliance, and Audit. Focus on building trust and enable widespread adoption.

*Trust. Safety. Transparency. Equality.*

## Small LLMs

Compact and efficient language models on resource-constrained compute. Domain specificity and knowledge embeddings.

*Domain knowledge AI everywhere*

## Q-Learning

Reinforcement learning based AI that learns from trial, error & rewards in an environment and learning optimal actions. Mastering complex tasks.

*Autonomous Learning & Adaptation*

## A.G.I

AI that has the (hypothetical) ability to understand, reason, & perform (and exceed) intellectual tasks typically associated with human intelligence

*Future of Human-Machine Collaboration*

# AI risk examples

How do we take action to drive both current and new AI systems towards being more responsible? To effectively achieve more Responsible AI, we need to understand challenges and risks around AI.

### 01. Data Integrity

Can predictions be corrupted by seemingly small or unintentional adversarial data changes?

### 02. Statistical Validity

Does the model measure what it was designed to measure?

### 03. Resiliency & Reliability

Can the AI scale while being architecturally resilient and reliable?

### 04. Fairness

Is inadvertent discrimination present based on gender, race, etc.?

### 05. Model Accuracy

How often does the model product the correct results?

### 06. Transparency

Are users enabled to understand how the AI was developed and how it generates its outputs?

### 07. Explainability

Does the company management understand and agree with how predictions are made? Can they explain how predictions are made?

### 08. Regulatory Compliance

# Global regulatory guidelines for AI are complex and evolving rapidly

## Canada

**Bill C-27, 2022** The Digital Charter Implementation Act includes the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act, and the Artificial Intelligence & Data Act (AIDA)
**GC Directive on Automated Decision-Making**
**Law 25** amending Quebec's **Act respecting the protection of personal information in the private sector** introduces new automated decision-making rules (s. 12.1) (Quebec).
**An Act to establish a legal framework for information technology** (Quebec, biometrics rules).

## Africa and Middle East

Kenya: Blockchain and AI taskforce (2018)
Tunisia: AI Task Force (2018)
South Africa: Sector specific initiative launched by Government for AI (2018)
Dubai – AI Ethics Principles and Guidelines (2018)

## Europe

**EU: EU Artificial Intelligence Act (2023),** Digital Services Act & Digital Markets Act
Finland: Released three reports 2017-19; last report focusses on ethics
Sweden: National Approach for AI (2018); launched national centre for AI innovation (2019)
Denmark: Strategy for Digital Growth (2018); National AI Strategy (2019)
The UK: AI Sector Deal in (2018)
Germany: National AI strategy (2018)
France: AI for humanity (2018)
Austria: Council on Robotics and AI (2017)
Spain: RDI Strategy in Artificial Intelligence (2019)
Italy: 'AI at the Service of Citizens' (2018); lab for AI created (2018)
Poland: 'Roundtable on AI strategy (2018)
Malta: Malta AI strategy Public Consultation (2019)
Estonia: Kraft Report (2019); AI taskforce (2018)
Netherlands: General Principles for the use of AI in Financial Sector (2019)

## US

**Biden Executive Order (2023)**
**NIST AI Risk Management Framework (2022)**
US AI Bill of Rights (2022)
State and Local policies
DOD AI Strategy (2019)
**NYC AI Hiring Act (2023)**

## Japan

Japan: AI Technology Strategy (2017) (part of Japan's Society 5.0 initiative); AI made a part of integrated innovation strategy (2018)

## Latin America

Mexico: 'Towards an AI Strategy in Mexico' white paper released (2018); no dedicated strategy yet; also has IA2030 Coalition that works with the government on AI
Brazil -E-Digital Strategy, digital transformation strategy addresses AI (2018);
Brazil, Argentina, Peru, Colombia, Costa Rica follow OECD principles on AI (2019)

## Australia

Publication: AI Ethics Framework Discussion Paper (2019)

## Asia

Singapore: Principles to promote FEAT in the use of AIDA in Singapore's financial sector (2018)
China – Beijing AI Principles Publication (2019)
Hong Kong – Ethical Accountability Framework Publication (2018)

## International

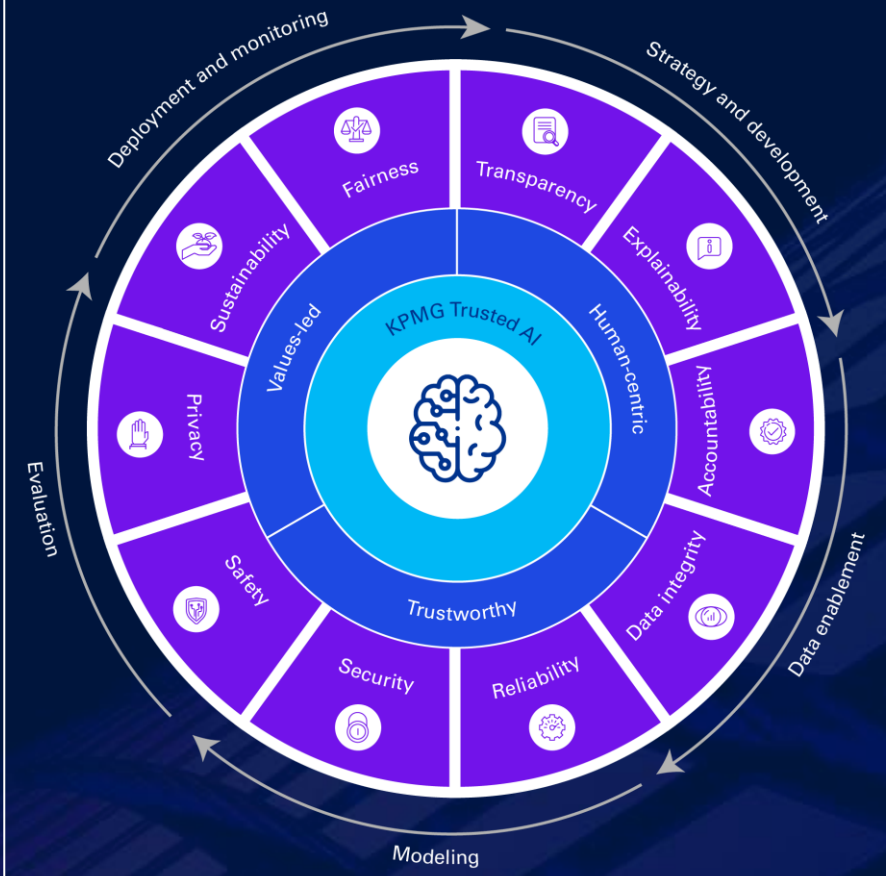
**OECD – The OECD council recommendations on Artificial Intelligence**. Global governance framework signed by 42 countries and non-OECD members Brazil, Argentina and Romania (2019)

# KPMG's Trusted AI Framework

# KPMG Trusted AI

We understand trustworthy and ethical AI is a complex business, regulatory, and technical challenge, and we are committed to helping clients put it into practice.

KPMG Trusted AI is our strategic approach and framework to designing, building, deploying and using AI solutions in a responsible and ethical manner so we can accelerate value with confidence.




## Values-driven

We implement AI as guided by our Values. They are our differentiator and shape a culture that is open, inclusive, and operates to the highest ethical standards.

## Human-centric

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.

## Trustworthy

We will strive to ensure our data acquisition, governance, and usage practices uphold ethical standards and comply with applicable privacy and data protection regulations, as well as any confidentiality requirements.

# The KPMG Trusted AI ethical pillars

**How we aim to deliver our commitments**

At KPMG, our approach to trusted AI rests on ten ethical pillars across the AI lifecycle:

## Fairness

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

## 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.

## 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 regulators 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.

## Security

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

## Safety

AI solutions should be designed and implemented to safeguard against harm to people, businesses, and property

## Privacy

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

## Sustainability

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

# Helping our clients

# What are clients asking for today?

Current Trusted AI pipeline and opportunities are centered around the following themes:

## Implement AI Governance

Organizations are seeking assistance from KPMG to either implement or redefine AI governance strategy and frameworks based on evolving regulatory landscape and AI specific risks.

## Establish AI Inventory

Organizations are struggling to get their hands around the AI models and applications deployed across their organizations, and seeking assistance from KPMG to help inventory and risk assess their AI applications.

## Regulatory Compliance

Driven by regulatory actions such as the EU AI Act, Whitehouse Exec Order, NY state law, and Colorado AI regulation, organizations are seeking assistance to ensure they have the necessary documentation, processes, controls, and procedures in place to meet the compliance requirements.

## Operationalizing AI Governance Programs

Organizations are seeking assistance from KPMG in operationalizing their AI security or governance frameworks, including policy development and procedural work.

## Use Case Development

Organizations are looking for support to identify and build specific AI applications in their environment and seeking assistance from KPMG to develop, train/test/tune, and deploy AI applications.

## AI Security

Organizations are seeking assistance to assess and develop AI security and privacy strategies; implement processes and tools to detect and respond to threats; implement AI security & privacy frameworks; and implement AI security tools..

## AI Assessments

Organizations are seeking assistance from KPMG to conduct AI model and application assessments, governance framework reviews, and risk and impact assessments across their AI landscape.  Clients are also seeking support to automate aspects of the assessment process across the AI lifecycle.

## Model validation and model risk management

Organizations are seeking support from KPMG to perform model validation, model risk management, update model vs non-model policy guidance, and interested in MRM as a service.

# A thoughtful roll-out of generative AI will allow you to simply address the associated risks

## Internal risks & considerations

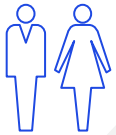
1. Breaking Confidentiality and Intellectual Property
2. Employee misuse and inaccuracies
3. Generative AI evolves
4. Talent Implications

## External risks & considerations

1. Misinformation, bias and discrimination
2. Copyright
3. Financial, brand and reputational risk
4. Cybersecurity
5. Adversarial attack

### Breaking Confidentiality and Intellectual Property

Many generative AI models are built to absorb user-inputted data to improve the model over time, and that could be used to **expose private or proprietary info**.



### Talent Implications

High-quality, **expert output can only be achieved with high-quality, expert queries**. Professionals need to be made aware that they're not just using a solution they're training and evolving it.

### Employee Misuse and Inaccuracies

The models generate responses based on input received, meaning there's a **risk they may provide false or malicious content**.

### Generative AI Evolves

As the world's understanding of AI evolve, we are already seeing a **rising number of global regulations**. It will continue to be integrated into many common applications.

### Misinformation, Bias and Discrimination

Generative AI can be used to create **deepfake images and videos**. These images and videos often look extremely realistic and lack forensic traces left behind in edited digital media.

### Copyright

Questions abound around **who owns content** once it's run through generative AI is difficult to answer.

### Cybersecurity

Cybercriminals can use Gen. AI to create more **realistic and sophisticated phishing scams** or credentials to hack into systems.

### Adversarial Attack

Even when trained to work within acceptable boundaries, Gen. AI models have proven to be vulnerable to **deliberate manipulation by sophisticated users**.

### Financial, Brand and Reputational Risk

If AI produced information were to be used into any deliverable, it **may constitute copyright or intellectual property infringement**. This could potentially cause your organization legal and reputational harm.

# Managing it All:

Managing risk associated with the design, development, deployment and management of AI solutions will require an understanding of each AI deployment; adapting legacy risk frameworks to embrace and incorporate emerging AI tools and trends; and adapting risk mindset with a focus toward monitoring outcomes, identifying model risk threats, and overall model risk management. To do this, the following are four pillars and representative actions Risk organizations should be focused on today:

## Understand AI Strategy and Roadmap

- Align current vision, strategy, and operating model for AI solutions
- Assess Board level oversight
- Inventory AI landscape within your organization, along with planned use cases, models, and tools.
- Ensure the use cases and vendor landscape for each AI solution are clearly understood
- Monitor 3rd party risks associated with data protection, storage of data, and access to confidential data
- Evaluate software tools that are being acquired to monitor ongoing data and AI pipeline security and privacy concerns (including poison and drift)
- Incorporate AI assessment into annual risk assessment process

## Compliance and Legal Risk

- Monitor AI regulatory developments
- Ensure appropriate stakeholder groups are implementing requirements and/or controls
- Align AI deployments and governance standards with appropriate regulatory guidelines and requirements
- Validate oversight of enterprise AI use and deployment standards
- Establish consistent contracting and AI deployment requirements for 3rd parties
- Ensure a mechanism has been established to identify, report, and manage AI vulnerabilities
- Assess ethical or societal impacts of planned AI usage
- Monitor legal considerations of external facing deployments

## Establish Governance

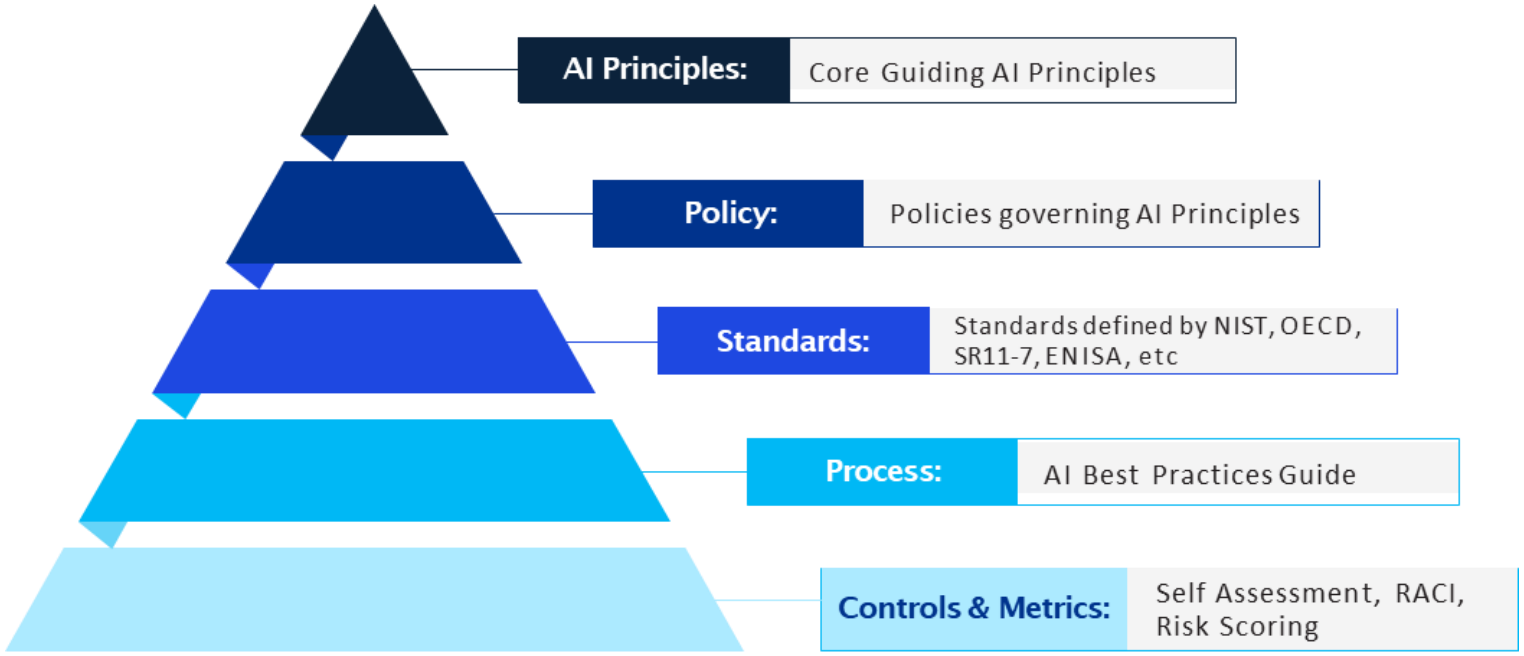
- Establish AI governance framework
- Develop policies that govern the use of AI throughout the organization with clearly defined roles and responsibilities
- Educate stakeholders on the use of AI, emerging risks around AI, and appropriate use policies
- Establish transparency principles and policies
- Incorporate AI into model risk management (MRM) framework including areas such as approved use, ongoing monitoring, and risk ratings
- Establish protocols for AI modeling usage, including business decisions vs experimental (Internal deployments),that align to MRM standards

## Monitor Usage and Deployments

- Perform AI risk assessments around areas such as compliance, governance, security, fairness, bias, accuracy, and explainability
- Assess access, API/interface, data security, privacy and change management controls specific to AI deployments
- Evaluate AI testing, training and deployment standards
- Assess financial reporting impact
- Identify KPIs to monitor AI outcomes, as well as detect anomalies, fraud, data poisoning
- Assess AI solution resiliency and reliability

# Establishing effective AI Governance

# AI Governance Considerations




- **Establish** your principles for AI that will guide your process in building the governance model and consider an enterprise-wide AI mission statement.
- **Reimagine** your existing governance model including your risk assessment process to uncover the risks of AI
- Ensure that your AI office is inclusive **diverse** group of **stakeholders** across Business, Technology, HR, Diversity amongst others.
- **Align** your AI deployments against appropriate standards and regulatory guidelines.
- **Monitor** your existing third and fourth parties to determine compliance against your responsible AI principles including existing low risk approved vendors.

# Lessons learned from building AI Governance models

01 Maintain security and privacy as core components of any governance model

02 Define a risk tiered governance approach

03 Develop and publicize a company wide Responsible Artificial Intelligence Charter

04 Help ensure that a diverse and representative group of stakeholders are involved in governance and model development

05 Re-evaluate your third-party risk management process

06 Reimagine your AI intake process

07 Build appropriate safeguards and measures to manage risks across the entire AI lifecycle, including ongoing monitoring

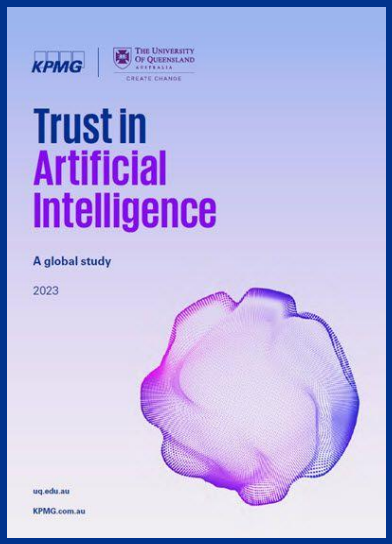
08 Align existing policies for AI

# Helpful Trusted AI Resources

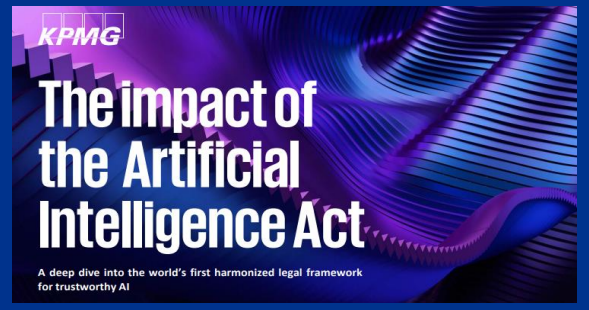
# Our Trusted AI Thought Leadership

Staying up-to-date on all things Trusted AI is no small feat, that is why we've collected some of our top Global thought leadership pieces around AI for you.









Additional resources:
KPMG Trusted AI