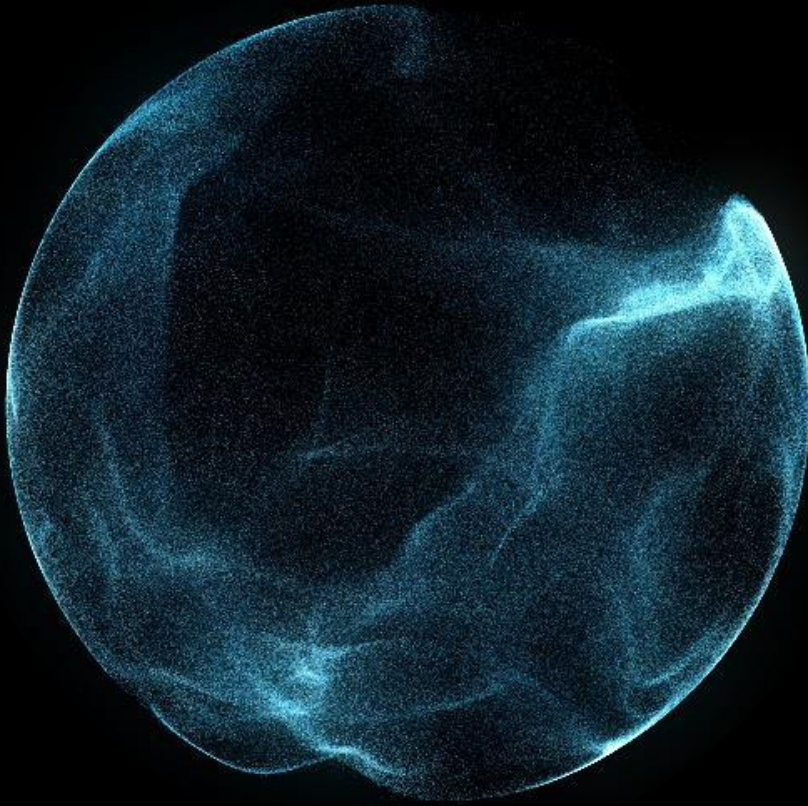# Deloitte.

The National Institute of Standards and Technology Artificial Intelligence Risk Management Framework (NIST AI RMF) emphasises the need for Trustworthy AI™

MAKING AN
IMPACT THAT
MATTERS
since 1845

2024

About the Deloitte AI Institute

The Deloitte AI Institute helps organisations connect all the different dimensions of the robust, highly dynamic and rapidly evolving AI ecosystem. The AI Institute leads conversations on applied AI innovation across industries, with cutting-edge insights, to promote human-machine collaboration in the "Age of With". Deloitte AI Institute aims to promote a dialogue and development of artificial intelligence, stimulate innovation, and examine challenges to AI implementation and ways to address them. The AI Institute collaborates with an ecosystem composed of academic research groups, start-ups, entrepreneurs, innovators, mature AI product leaders, and AI visionaries, to explore key areas of artificial intelligence including risks, policies, ethics, future of work and talent, and applied AI use cases.

Combined with Deloitte's deep knowledge and experience in artificial intelligence applications, the Institute helps make sense of this complex ecosystem, and as a result, deliver impactful perspectives to help organisations succeed by making informed AI decisions.

No matter what stage of the AI journey you're in; whether you're a board member or a C-suite leader driving strategy for your organization, or a hands-on data scientist, bringing an AI strategy to life, the Deloitte AI institute can help you learn more about how enterprises across the world are leveraging AI for a competitive advantage. Visit us at the Deloitte AI Institute for a full body of our work, subscribe to our podcasts and newsletter, and join us at our meet ups and live events. Let's explore the future of AI together.

www.deloitte.com/us/AIInstitut

The NIST AI RMF sets the stage for future regulations and provides organisations with a roadmap to adapt risk management for AI

The National Institute of Standards and Technology Artificial Intelligence Risk Management Framework[1] (NIST AI RMF) advances prior guidance set forth to aid organisations in understanding, assessing and managing AI risk, and provide trust in the evolving technological landscape[2, 3, 4, 5]. As organisations expand their use of AI and other automated systems to help realize efficiencies and technology-enable processes, regulators[2, 3, 4] continue to refine guidance[2, 3, 4, 5] in their efforts to safeguard the public.

In January 2023, the U.S. Department of Commerce, through NIST, published the first edition of its AI RMF and accompanying playbook to provide guidance for organisations on the creation, deployment and continual management of trustworthy and ethical AI systems. As the AI technological landscape continues to evolve, NIST intends to further develop additional guidance[1], metrics and methodologies. Though voluntary, the AI RMF provides insights into potential future regulation and actionable steps for organisations to integrate AI risk management into their existing enterprise risk management (ERM) practices.

Organisations across the AI implementation maturity spectrum can leverage the AI RMF as a resource in designing, developing, deploying or utilising AI technology to establish trustworthy AI and help mitigate AI risks.

The NIST framework provides a socio-technical perspective on AI risk management, aligning trustworthy AI technologies to organisational purpose and values. Organising around these core functions (figure 1) enables transparency and alignment between disparate groups of AI stakeholders, including developers, data scientists, operational users and management. As organisations implement and evolve their use of AI, the NIST AI RMF can serve as a roadmap to effectively integrate AI risk management.

Figure 1 | Definition of core functions govern, map, measure and manage according to the NIST AI RMF.

Govern
Organise people, process, and structures to create policies, accountabilities and culture around AI risk

Map
Establish the context to identify and frame organisational risks related to AI tools

Measure
Employ tools and methodologies to monitor, track and analyse AI risks and related impacts

Manage
Prioritise and control AI risks in line with enterprise risk management practices

[3]

NIST AI RMF and the journey to trustworthy AI

Why is there a need for proactive AI risk management?

- The explosion of sophisticated AI technologies has led to an increase in applications in organisations and rapidly expanded the risk footprint associated with AI

- Emerging regulations[2, 3, 4, 5] will require organisations to be able to better explain how their AI models work and how they adhere to privacy standards

- Integration of AI technologies into existing business processes increases the need for transparency within AI models to determine compliance with broader regulatory requirements on quality and fairness.

The NIST AI RMF reinforces the need to integrate AI risk management into broader enterprise risk management practices. The framework emphasizes consideration of AI-specific nuances when considering organisational risk tolerance, prioritisation and integration, identifying overlapping risks such as data privacy concerns or environmental implications from computing power. The framework is a call to action for organisations to **establish and maintain appropriate responsibilities and accountabilities to enable effective risk management.**

AI risk management activities should be embedded across the AI lifecycle by stakeholders across the organisation. Trustworthy AI depends upon continuous and timely design, testing, evaluation, verification and validation of AI systems and associated risks. The responsibilities for these activities will vary throughout the AI lifecycle (figure 2) and should be clearly defined to enable effective risk governance.

AI risk challenges:

- Aligning risk metrics with third-party technologies

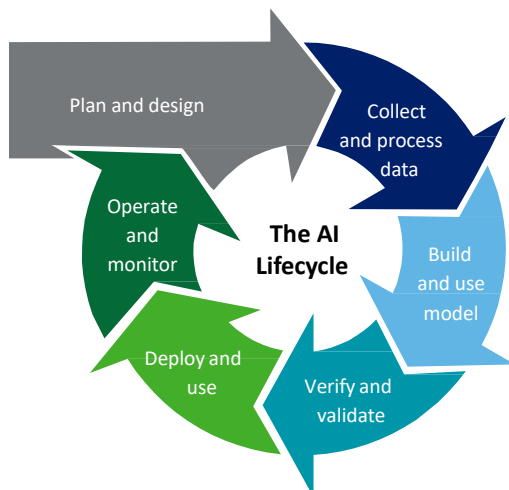- Opaqueness within AI systems and underlying algorithms

- Differences between controlled and real-world environments

- Establishing human baselines, review and oversight

Figure 2 | The AI lifecycle, adapted from the NIST AI RMF

The AI Lifecycle: Plan and design → Collect and process data → Build and use model → Verify and validate → Deploy and use → Operate and monitor

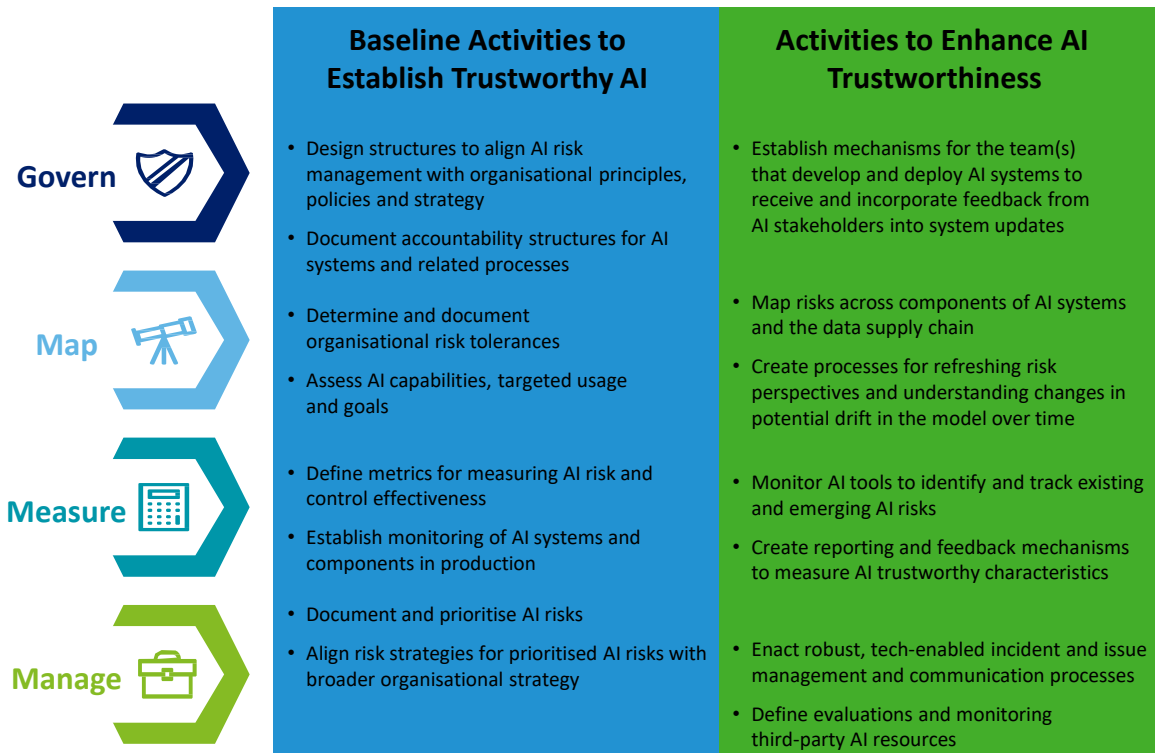## How can organisations get started?

To effectively leverage the NIST AI RMF, organisations should begin by assessing their current AI capabilities and strategy, as well as how it intersects with broader ERM efforts. The framework is intended to be flexible, helping enable organisations to align practices with applicable laws, regulations and norms which may differ by industry or sector, as illustrated in Figure 3.

Once an effective baseline is established, organisations can start to apply framework insights on measuring risks, risk tolerance, risk prioritisation and integration of risk management concepts related to AI.

As organisations' AI capabilities mature, the NIST framework and its core functions should be revisited and supporting risk management capabilities should continue to be iterated upon to strengthen trustworthy AI.

**Figure 3** | Illustrative activities aligned to the NIST AI RMF core functions

	Baseline Activities to Establish Trustworthy AI	Activities to Enhance AI Trustworthiness
Govern	• Design structures to align AI risk management with organisational principles, policies and strategy • Document accountability structures for AI systems and related processes	• Establish mechanisms for the team(s) that develop and deploy AI systems to receive and incorporate feedback from AI stakeholders into system updates
Map	• Determine and document organisational risk tolerances • Assess AI capabilities, targeted usage and goals	• Map risks across components of AI systems and the data supply chain • Create processes for refreshing risk perspectives and understanding changes in potential drift in the model over time
Measure	• Define metrics for measuring AI risk and control effectiveness • Establish monitoring of AI systems and components in production	• Monitor AI tools to identify and track existing and emerging AI risks • Create reporting and feedback mechanisms to measure AI trustworthy characteristics
Manage	• Document and prioritise AI risks • Align risk strategies for prioritised AI risks with broader organisational strategy	• Enact robust, tech-enabled incident and issue management and communication processes • Define evaluations and monitoring third-party AI resources

How the essential characteristics of the NIST AI RMF align with Deloitte's Trustworthy AI Framework™

As AI and other advanced automated systems are becoming increasingly common tools used by organisations, Deloitte recognised the need to approach these evolving technologies in an ethical and responsible manner. As pictured in figure 4, **Deloitte's Trustworthy AI Framework has empowered organisations to build trustworthy AI and helped prepare them for the growing regulatory focus on AI and other automated systems.**

The NIST AI RMF outlines seven characteristics for achieving responsible use of AI systems and effectively managing AI risk: **valid and reliable, accountable and transparent, safe, secure and resilient, explainable and interpretable, privacy-enhanced and fair.** Each characteristic builds upon the socio-technical viewpoint NIST advocates when implementing and managing AI technologies, however the accountability and transparency attributes also pertain to the external processes and context surrounding the AI systems.

Figure 4 | Deloitte's Trustworthy AI Framework
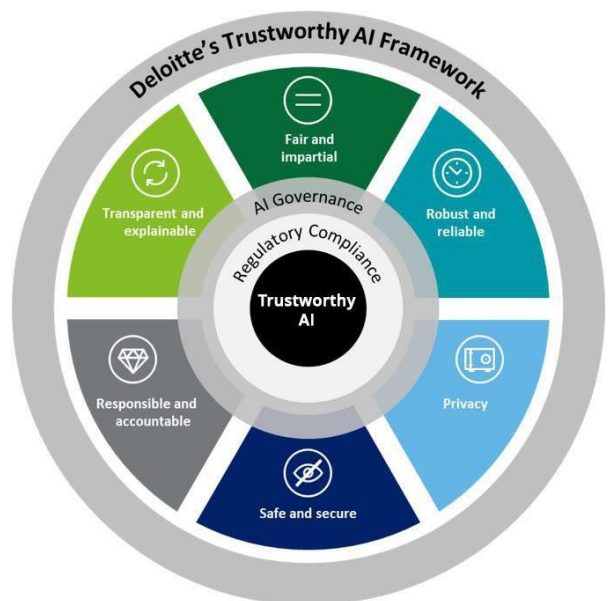
Over the past decade, Deloitte developed its Trustworthy AI Framework based on hands-on experience and cross-industry leading practices to help clients throughout the AI lifecycle manage AI risk. Deloitte's Trustworthy AI Framework is comprised of six characteristics: **fair and impartial, robust and reliable, privacy, safe and secure, responsible and accountable and transparent and explainable**.
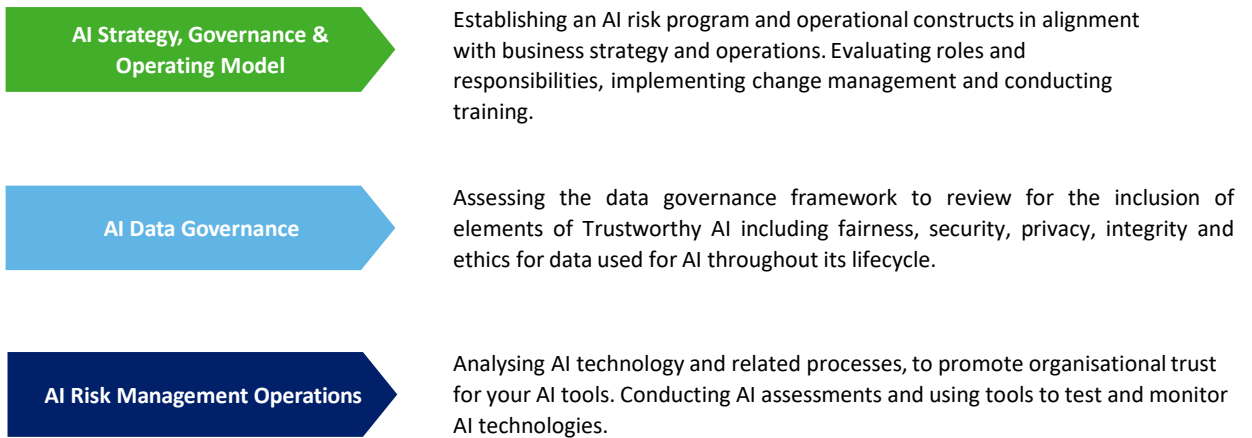
The characteristics outlined by the NIST AI RMF align well with Deloitte's Trustworthy AI Framework, as highlighted in figure 5 below, and the focus on trustworthiness can help organisations to effectively utilise a variety of automated systems while feeling confident in the security and performance of their AI models.

Figure 5 | NIST Characteristics of trustworthy AI and how they map to Deloitte's Trustworthy AI Framework

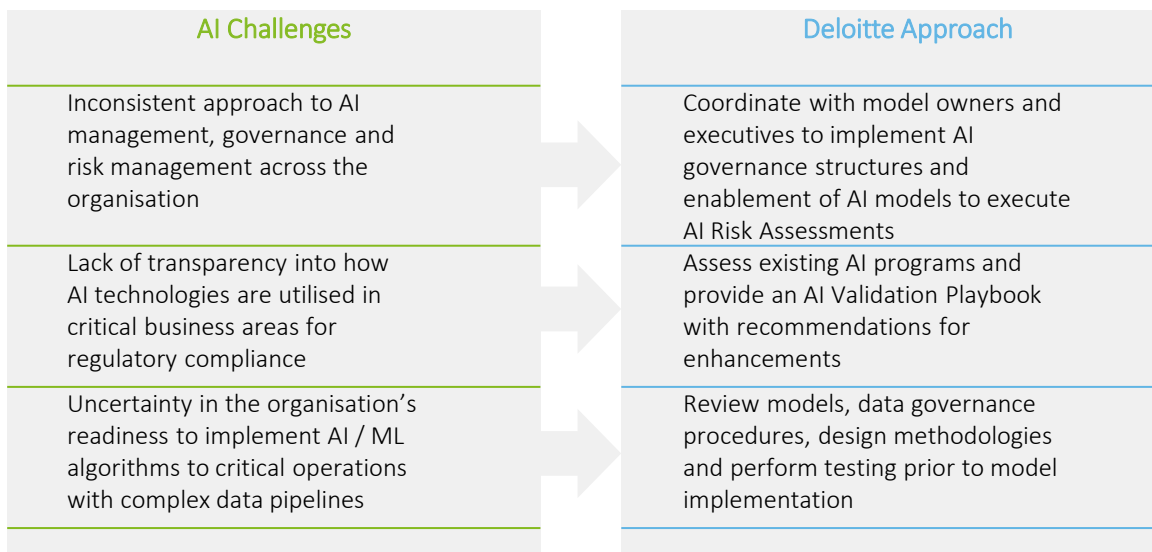| NIST Artificial Intelligence Risk Management Framework		
NIST Characteristics of Trustworthy AI	**NIST Description**	**Deloitte Trustworthy AI Framework**
Valid & Reliable	Confirms the tool is fit for its intended purpose and can perform its requirements	✓ Robust & Reliable
Accountable & Transparent	Provides that access is given to appropriate personnel as required to understand and resolve issues to safeguard information	✓ Transparent & Explainable, Responsible & Accountable
Safe	Protects life, health, property and the environment through responsible practices surrounding design, deployment, decision-making and documentation	✓ Safe & Secure
Secure & Resilient	Protects against adverse or unexpected environmental changes or use, and maintains confidentiality, integrity, and availability	✓ Safe & Secure, Robust & Reliable
Explainable & Interpretable	Provides those overseeing AI systems with deeper insights into the functionality, trustworthiness and outputs of the system	✓ Transparent & Explainable, Responsible & Accountable
Privacy-Enhanced	Protects human autonomy, identity and dignity, including values such as anonymity and confidentiality	✓ Privacy
Fair – With Harmful Bias Managed	Protects against systemic, computational/statistical and human-cognitive biases to promote transparency and fairness in outputs	✓ Fair & Impartial

How Deloitte Can Help

At Deloitte, we utilise demonstrated approaches and tactics for operationalising Deloitte's Trustworthy AI Framework through foundational risk management practices across people, processes and technologies to automate and continuously monitor risk posture. As we help our clients on the journey to Trustworthy AI, Deloitte uses multiples approaches including:

AI Strategy, Governance & Operating Model

Establishing an AI risk program and operational constructs in alignment with business strategy and operations. Evaluating roles and responsibilities, implementing change management and conducting training.

AI Data Governance

Assessing the data governance framework to review for the inclusion of elements of Trustworthy AI including fairness, security, privacy, integrity and ethics for data used for AI throughout its lifecycle.

AI Risk Management Operations

Analysing AI technology and related processes, to promote organisational trust for your AI tools. Conducting AI assessments and using tools to test and monitor AI technologies.

Deloitte has served as a trusted adviser in assisting clients establish leading risk management and governance processes and continues to do so as our clients grapple with the risks of AI technologies (figure 6). utilising our Trustworthy AI framework and deep knowledge of cross-industry AI guidance, we are dedicated to assisting clients through the changing regulatory landscape and effectively identify, plan for and manage AI risk.

Figure 6 | Applying Deloitte's Trustworthy AI Framework

AI Challenges	Deloitte Approach
Inconsistent approach to AI management, governance and risk management across the organisation	Coordinate with model owners and executives to implement AI governance structures and enablement of AI models to execute AI Risk Assessments
Lack of transparency into how AI technologies are utilised in critical business areas for regulatory compliance	Assess existing AI programs and provide an AI Validation Playbook with recommendations for enhancements
Uncertainty in the organisation's readiness to implement AI / ML algorithms to critical operations with complex data pipelines	Review models, data governance procedures, design methodologies and perform testing prior to model implementation

Let's start a conversation

Samresh Ramjith
Risk Advisory Africa
Cyber Leader
Mobile: +27 11 517 4047
sramjith@deloitte.co.za

Jania Okwechime
Consulting West Africa
AI & Data Analytics Partner
Mobile: +23 41 904 1778
jokwechime@deloitte.com.ng

Leishen Pillay
Risk Advisory Africa
Privacy and AI
Governance Partner
Mobile : +27 11 209 6418
lpillay@deloitte.co.za

Quentin Williams
Consulting Africa
AI Leader
Mobile: +27 11 304 5786
qwilliams@deloitte.co.za

Risk Advisory Africa Leaders

Greg Rammego
Risk Advisory Africa
Managing Director
Mobile: +27 11 806 5255
grammego@deloitte.co.za

Temitope Aladenusi
Risk Advisory
West Africa Leader
Mobile: +234 19 041 730
taladenusi@deloitte.com.ng

Urvi Patel
Risk Advisory Africa
East Africa Leader
Mobile: +254 71 405 6887
ubpatel@deloitte.co.ke

Tricha Simon
Risk Advisory
Central Africa Leader
Mobile: +263 867 700 0261
tsimon@deloitte.co.zm

Endnotes

1.   National Institute of Standards and Technology AI Risk Management Framework

2.   The AI Bill of Rights follows the Executive Order 13960: Promoting the Use   of Trustworthy Artificial Intelligence in the Federal Government (December   2020)

3.   Executive Order 13859: Maintaining American Leadership in Artificial   Intelligence (February 2019)

4.   Office of Management and Budget (OMB)   Memorandum M-21-06: Guidance for Regulation of Artificial Intelligence Applications (November 2020), White House Office of Science and Technology   Policy (OSTP): American AI Initiative: Year One Annual Report (February 2020)

5.   International initiatives include the Organisation for Economic Co-operation and Development (OECD): 2019   Recommendations on Artificial Intelligence, and the European Union Artificial   Intelligence Act proposal (April 2021).

Deloitte.