Deloitte.

Executive Summary for Trustworthy AI™

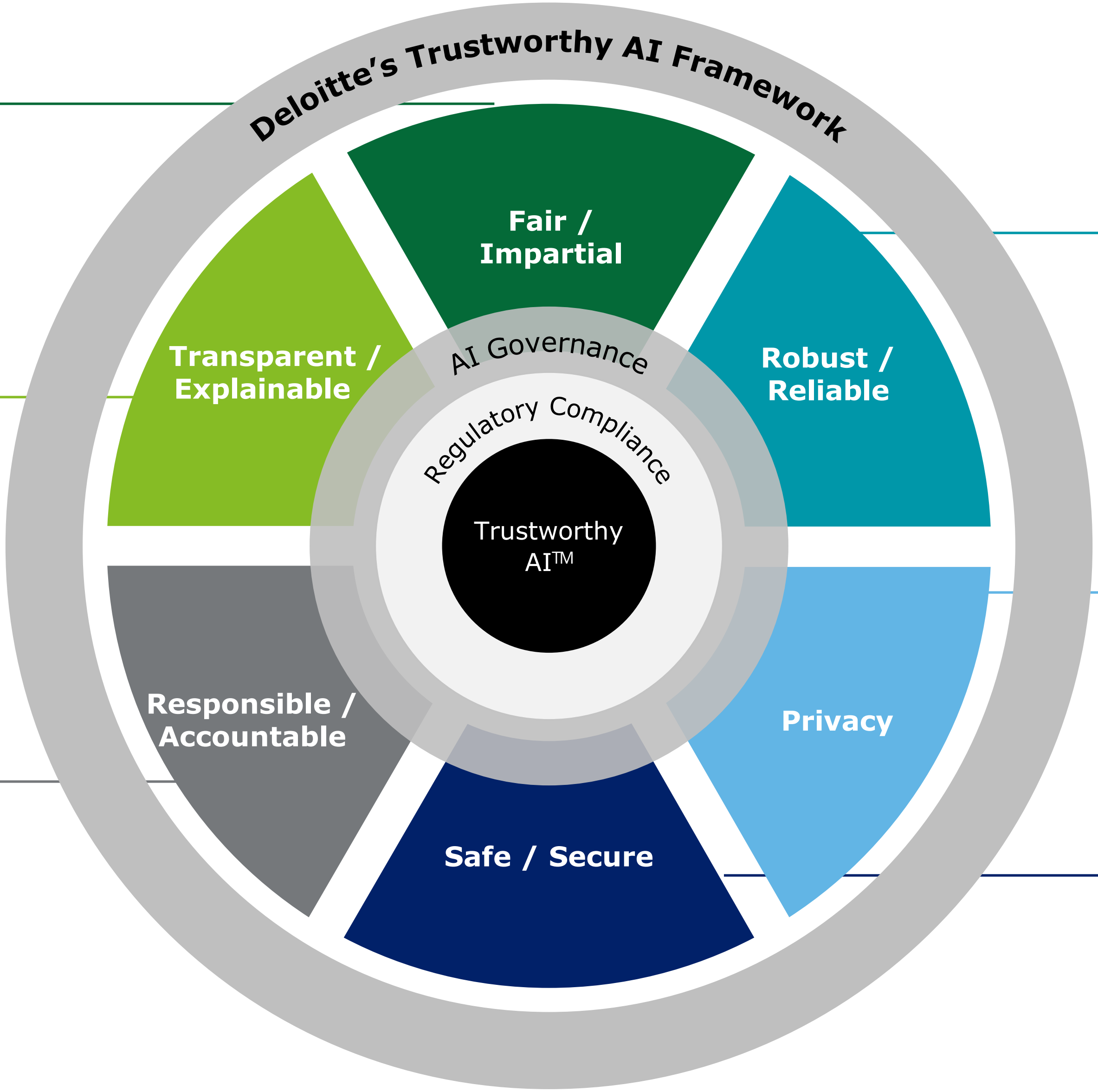
Principles and Examples

Applying Deloitte’s six-part framework is an effective first step in diagnosing the ethical health of AI while maintaining customer privacy and abiding by relevant policies

Fair / Impartial
AI applications include internal and external checks to help enable equitable application across all participants

Transparent / Explainable
All participants are able to understand how their data is being used and how AI systems make decisions; algorithms, attributes, and correlations are open to inspection

Responsible / Accountable
Policies are in place to determine who is held responsible for the output of AI system decisions



Robust / Reliable
AI systems have the ability to learn from humans and other systems and produce consistent and reliable outputs

Privacy
Consumer privacy is respected and customer data is not used beyond its intended and stated use; consumers are able to opt in/ out of sharing their data

Safe / Secure
AI systems can be protected from risks (including cyber risks) that may cause physical and/or digital harm

Trustworthy AI™ Framework – Principles and Examples (1)

Pillar	Key Issues	Key Steps to Address Issues
AI Governance and Regulatory Compliance *Organizations will need to develop effective governance and regulatory compliance over the design, deployment, and operational phases of AI to help safeguard ethics and build trustworthy AI*	Organizations need to identify AI-related risks and strike an appropriate balance between AI enablement and risk management Organizations need to implement proactive regulatory engagement plans to address the policy and legal uncertainties of deploying AI and to monitor ongoing regulatory and legal developments Organizations need to have the right leadership and resources (with appropriate skills) to develop, implement, and monitor AI systems	Develop an overall policy which defines the purpose and objectives of the organization's use of AI and how AI is to be used. The development of this policy should consider the wider impact of the organization's use of AI on external stakeholders Perform a risk assessment to identify risks related to the application of the organization's AI (evaluating each dimension of the Trustworthy AI Framework) Informed by a risk assessment, design and implement control activities that mitigate risks pertaining to each of the dimensions of the Trustworthy AI framework and data management/integrity Establish a process to regularly evaluate whether the use of AI is appropriate in the circumstances, whether AI is yielding desired benefits, and make any adjustments as appropriate
Fair / Not Biased *AI must be designed and trained to follow a fair, consistent process and make fair decisions. It must also include internal and external checks to reduce discriminatory bias*	Organizations need to determine what "fairness" means for their use AI Organizations need to address concerns from stakeholders regarding biased results from AI	Evaluate if there is a significant level of risk related to bias associated with the organization's use of AI Consider purpose of AI use, input from stakeholders, organization policies, and relevant regulations to determine benchmarks used to measure and evaluate "fairness" Evaluate and monitor data used by AI to minimize or mitigate bias

Trustworthy AI™ Framework – Principles and Examples (2)

Pillar	Key Issues	Key Steps to Address Issues
Transparent & Explainable *Participants have a right to understand how their data is being used and how the AI is making decisions. The AI's algorithms, attributes, and correlations must be open to inspection, and its decisions must be fully explainable*	Organizations need to consider and address stakeholders' transparency expectations regarding use of AI as well as any regulatory requirements	Consider transparency expectations or regulations and determine if use of algorithms -- "black-box" vs. "non black-box" models -- are appropriate Determine what information regarding use and functionality of AI should be obtained, tested, documented, and communicated internally and externally Implement processes to obtain, retain, modify, and disseminate such information on a consistent basis
Responsible & Accountable *Organizations using AI need to have policies that clearly establish who is responsible and accountable for AI systems output*	Organizations need to consider if their use of AI is socially responsible Stakeholders need to understand responsibilities regarding use of AI and be held accountable (through structured mechanisms) to those responsibilities	Consider organization's mission, purpose of AI use, and impact of AI use internally and externally to evaluate if AI use is socially responsible and make adjustments as necessary Establish clear operating structures and reporting lines to provide appropriate oversight of the organization's use of AI and an established protocol to follow if something goes wrong
Robust & Reliable *AI systems must generate consistent and reliable outputs and scale up well. If it fails, it must fail in a predictable, expected manner*	AI systems need to perform as intended in less than ideal conditions and when encountering unexpected situations and data Organizations need to sufficiently monitor its AI systems and address deficiencies in a time-sensitive manner	Develop a testing regime that includes variability (e.g., changes in the system or training data) to evaluate if AI is robust enough to function as intended despite differences in the environment Establish measures for reliability and consistency Monitor data inputs or training data, AI code, and related outputs. Investigate and resolve exceptions, discrepancies, unintended outcomes in a timely manner

Trustworthy AI™ Framework – Principles and Examples (3)

Pillar	Key Issues	Key Steps to Address Issues
Privacy *AI systems must comply with data regulations and only use data for the stated and agreed-upon purposes*	Organizations need to know what data is being collected and why and, if applicable, provide customers with an appropriate level of control over their data Organizations need to consider whether the AI systems in use generate trust with customers, employees, and other stakeholders and how to retain that trust Organizations need to identify, evaluate, and monitor compliance with data privacy rules and regulations	Determine what information to collect, how long the information will be used, how the information will be retained, how consent to use data will be obtained, how information will be used, and how information will be disposed of Develop policies and processes over securely retaining data (e.g., encryption), disposing of data, obtaining consent as appropriate, communicating what is obtained, how it is used, and how it is maintained and disposed. Enforce accountability through actions.
Safe/Secure *AI systems must be protected from risks that might lead to physical and/or digital harm*	Organizations need to consider whether AI helps to maintain or increase safety and security Organizations' cyber-infrastructure and expertise need to be robust enough to tackle AI-specific cyber risks (i.e. adversarial manipulation of AI models, reverse engineering of data) Organizations need to have strategies to achieve employee awareness of AI risks	Determine security objectives and risks related to use of AI based on variation of AI being utilized, how it will be used, requirements related to security, and interactions/reliance with external parties Establish testing processes as well as incremental preventative and detective/monitoring controls, beyond traditional IT asset management controls, specific to the safety and security of AI's underlying technology Communicate AI safety/security principles internally and externally so that employees and others understand expectations

This presentation contains general information only and Deloitte is not, by means of this presentation, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This presentation is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor.

Deloitte shall not be responsible for any loss sustained by any person who relies on this presentation.

About Deloitte

As used in this document, 'Deloitte' means Deloitte & Touche LLP, which provides audit, assurance, and risk and financial advisory services and Deloitte Consulting LLP, which provides strategy, operations, technology, systems, outsourcing and human capital consulting services. These entities are separate subsidiaries of Deloitte LLP. Please see www.deloitte.com/us/about for a detailed description of our legal structure. Certain services may not be available to attest clients under the rules and regulations of public accounting.