

Deloitte.



Unlocking The Potential of AI
Through Trust

2023

Introduction

Change of paradigm, a tool or a threat to humanity? The true societal impact of Artificial Intelligence (AI) systems is yet to be fully realised. However, many already see AI as an engine for productivity and economic growth. As organisations compete to be the first to unlock and realise the full potential of AI for their organisations, governments and regulators across the world have started the hard task of putting legislation and regulatory frameworks around a technology that is constantly evolving. While there is still a lot of uncertainty around the risks due to AI technologies, there needs to be some caution displayed to truly understand these, particularly where risks and harms to individuals may arise. In addition, privacy and security concerns are still the leading causes that are limiting investments in AI-based solutions. However, with the current buzz around AI, even an organisation that is not currently considering it will be inclined to do so as the technologies evolve and mature. From this perspective, it is important to start thinking about AI use cases for your business and be ready to implement such solutions in a manner that builds customer confidence and aligns with the regulatory requirements. There is no doubt that companies that have an issue with how and where they deploy AI technologies will suffer from significant reputational damage.

In this article, we will look at the AI Risk Landscape and how risk mitigation can be achieved through taking a trustworthy AI approach.



AI Risk Landscape

As AI technologies have continued to evolve, so have the risks resulting from the use of these technologies. It is crucial to be cognizant of these risks as they can have a significant impact on users as well as the organisations that implement such technologies.

While the AI risk landscape is constantly evolving, some key risks due to AI systems can be broadly classified as follows:



- Lack of Transparency – This is one of the biggest challenges that organisations need to overcome. While keen to progress and even understand the AI journey themselves, a lack of transparency around the impacts on users/consumers may cause unease with the user/customer base. This challenge is compounded when AI systems are not clear or understandable to the people who will be using them or are impacted by the decision making of the AI system. Any kind of automated decision making, for example, whether you qualify for a loan or not, must be explained and the decision approach transparent. In addition, it must have the ability to be reviewed and overridden manually with that learning feeding back into the technology.
- Bias and Discrimination – There are two main types of bias that can occur in AI systems. One is centred around algorithmic or data bias which may occur when the data sets being used to train the algorithms are themselves biased, and the second can occur as a societal AI bias, where the AI reflects social intolerance or institutional discrimination. As a result, there is a risk that when an AI system treats some people or groups differently or unfairly based on factors that are irrelevant to the task or decision, this can often result in negative outcomes for those people or groups.
- Inference Attacks - These are techniques used for analysing data (such as input or output of the AI system) to try and infer or deduce more information about the data to which the attacker does not have access to. For example, an inference attack might try to figure out what kind of information was used to train an AI system, how this could be corrupted or what kind of information is stored in a database that the AI system accesses.

Unlocking The Potential of AI Through Trust

- Privacy Concerns - According to the [European Consumer Organisation in 2020](#), a survey showed that 45-60% of Europeans agree that AI will lead to more abuse of personal data. The relationship between data privacy and artificial intelligence is quite nuanced. AI algorithms may unavoidably require private data; however, the use of such data can result in consumers feeling that the processing is invasive or that their data is insecure. A basic and core principle of the General Data Protection Regulation (GDPR) is that of data minimisation. Organisations tend to do this through the deidentification of data in some form or another (anonymisation, pseudonymisation, data segregation) and subsequent technical and organisational measures to secure this data. AI raises a risk of reidentification, known in the privacy world as the Mosaic Effect. This can occur when individual data sets which are non-identifiable on their own are combined resulting in reidentification.
- IP Infringement - AI systems are trained on large archives of data to recover patterns and relationships. These forms of data may include information that is owned by other individuals or organisations. As such, outputs generated by these AI systems may infringe upon the Intellectual Property of others. We are already witnessing these risks being realised as multiple AI platforms are facing lawsuits regarding the unlicensed use of protected data.
- Mis- and Disinformation Harms – There is a real risk that generative AI will strengthen the current misinformation problem and overburden already stretched content moderation systems. This may lead to a loss of consumer trust and can also have severe societal harms by over reliance on content that is less than accurate and giving power to those who might strategically wish to manipulate information.
- Human-Computer Interaction Harms - As AI-based systems advance, interactions with systems can seem more like interactions with other humans. Users may be unaware when they are engaging with AI systems and these interactions can lead to unsafe use as users may overestimate the outputs, and this may create new avenues to exploit and violate user privacy.



Risk Mitigation Through Trustworthy AI



While the risks of AI technology do exist, there is also no doubt about the benefits that can and will be realised. However, the social and economic opportunities of AI may not be appropriately realised if the public's concerns about the risks of AI outweigh their perception of the benefits. It is therefore crucial to ensure that AI technologies evolve and are deployed in ways that consumers and users can reasonably trust.

Trustworthy AI, also known as ethical or responsible AI, has been proposed as a means for mitigating the risks and enhancing consumer/user trust in such systems. This is an umbrella term that consolidates a number of components which, according to the independent high-level expert group on AI established by the European Commission, consist of the premise that Trustworthy AI must be:

lawful, respecting all applicable laws and regulations;

ethical, respecting ethical principles and values; and

robust, from a technical perspective but also considering the social environment.

Keeping the above in mind, applying a human centric trustworthy AI by design approach will go a long way towards propelling innovative AI efforts, while being cognizant of the risks that need to be mitigated.

Deloitte's Six Dimensions for Trustworthy AI

Fair and Impartial

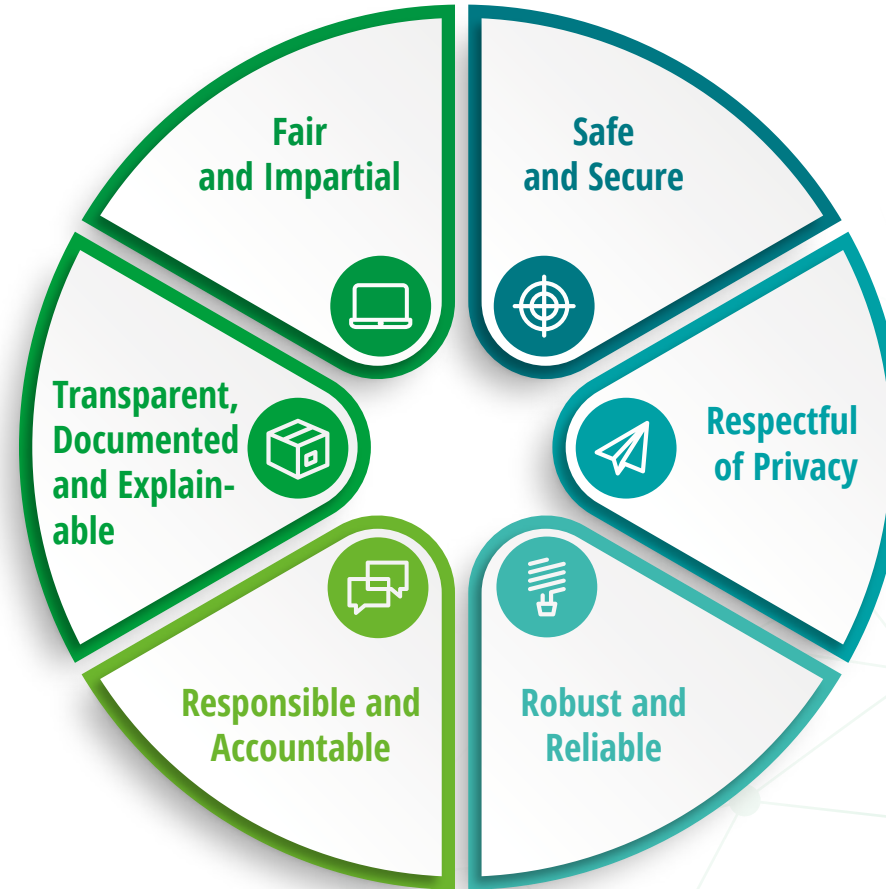
AI systems should make decisions that follow a consistent process and apply rules fairly, as well as incorporate internal and external checks to remove biases that might lead to discriminatory or differential outcomes, to help ensure results that are not merely technically correct but considerate of the social good.

Transparent, Documented and Explainable

AI systems may not operate as "black boxes"; all parties engaging with an AI should be informed that they are doing so and be able to inquire as to how and why the system is making decisions.

Responsible and Accountable

The increasing complexity and autonomy of AI systems may obscure the ultimate responsibility and accountability of companies and human beings behind the decisions and actions of these systems; policies should be in place to clearly assign liability and help ensure that parties impacted by AI can seek appropriate recourse.



Safe and Secure

Just as we currently depend on the consistent performance of human professionals to help ensure that our daily activities are safe and healthy, we should be able to depend on equivalent or even greater reliability as we enable more of our systems with AI.

Respectful of Privacy

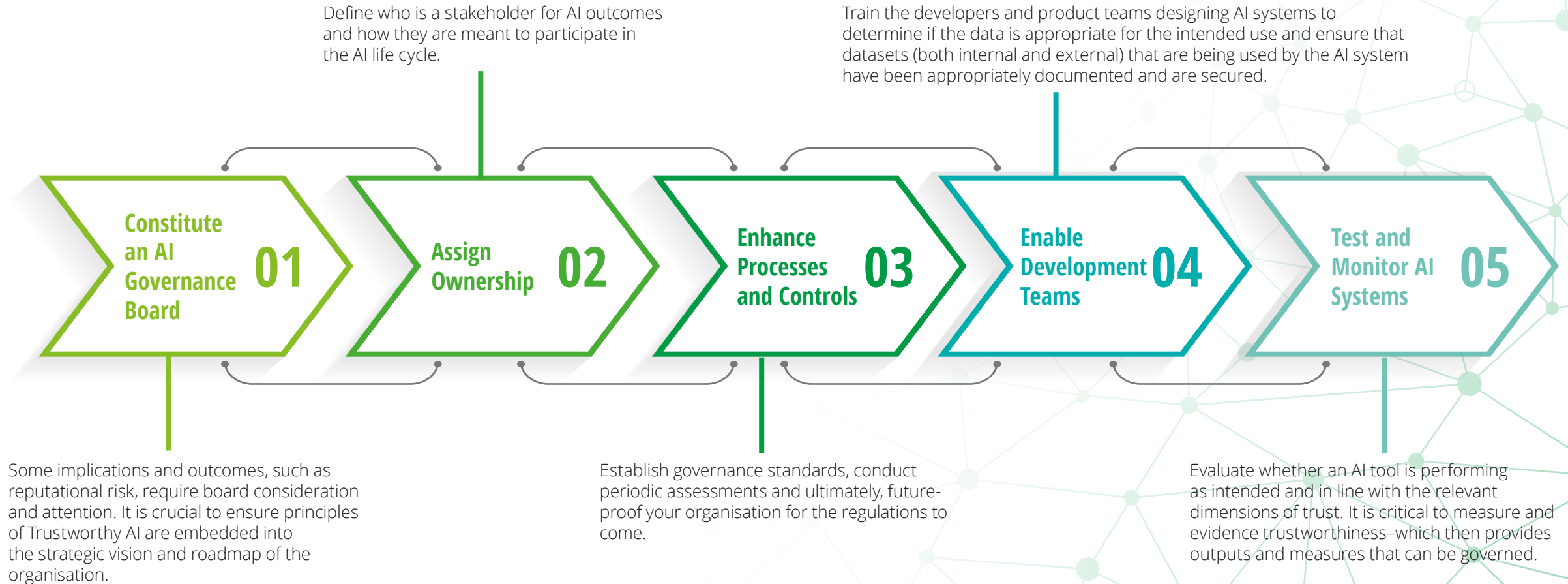
As AI systems often rely on gathering large amounts of data to effectively accomplish their tasks, we should ensure that all data is gathered appropriately and with full awareness and consent, and then securely deleted or otherwise protected from further, unanticipated use.

Robust and Reliable

As AI systems take greater control over more critical processes, the danger of cyberattacks and other harms expands significantly. Appropriate security measures should be put in place to help ensure the integrity and safety of the data and algorithms that drive AI.

Implementing Trustworthy AI

Trustworthy AI requires a careful assessment of risks, their impact and the value obtained from AI systems. A strong AI governance ecosystem is key for promoting Trustworthy AI across the organisation. Some key steps organisations can undertake to enable Trustworthy AI are as follows:



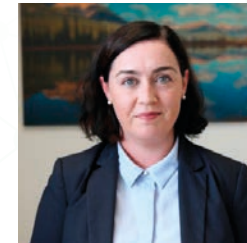
How Can Deloitte Help?

We can collaborate with you to develop and implement a trustworthy AI framework suitable to your business needs. We take a risk-based approach that is specifically aligned with your business strategy, industry needs, regulatory landscape, and will ensure you are ready for the present and future regulatory requirements.



Colm McDonnell
Partner

Head of Technology, Media
& Telecommunications
Deloitte Ireland



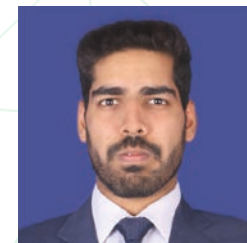
Nicola Flannery
Director

Data Privacy
& Internet Regulation
Deloitte Ireland



Sophie Vinti
Senior Manager

Internet Regulation Lead,
SME in Trust by Design
Deloitte Ireland



Ravin Nandle
Manager

Trustworthy AI
and Data Privacy
Deloitte Ireland



Important notice

At Deloitte, we make an impact that matters for our clients, our people, our profession, and in the wider society by delivering the solutions and insights they need to address their most complex business challenges. As the largest global professional services and consulting network, with over 312,000 professionals in more than 150 countries, we bring worldclass capabilities and high-quality services to our clients. In Ireland, Deloitte has over 3,000 people providing audit, tax, consulting, and corporate finance services to public and private clients spanning multiple industries. Our people have the leadership capabilities, experience and insight to collaborate with clients so they can move forward with confidence.

This publication has been written in general terms and we recommend that you obtain professional advice before acting or refraining from action on any of the contents of this publication. Deloitte Ireland LLP accepts no liability for any loss occasioned to any person acting or refraining from action as a result of any material in this publication.

Deloitte Ireland LLP is a limited liability partnership registered in Northern Ireland with registered number NC1499 and its registered office at 19 Bedford Street, Belfast BT2 7EJ, Northern Ireland.

Deloitte Ireland LLP is the Ireland affiliate of Deloitte NSE LLP, a member firm of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee ("DTTL"). DTTL and each of its member firms are legally separate and independent entities. DTTL and Deloitte NSE LLP do not provide services to clients. Please see www.deloitte.com/about to learn more about our global network of member firms.

© 2023 Deloitte Ireland LLP. All rights reserved.

Designed by CoRe Creative Services. RITM1527564