



# Responsible and Explainable AI

Digital Trust



The use of data and algorithms to automate processes and drive decision-making in business is leading to greater efficiency and accuracy. Machines are becoming smarter and more independent from algorithms that can learn and make predictions from structured and unstructured data.

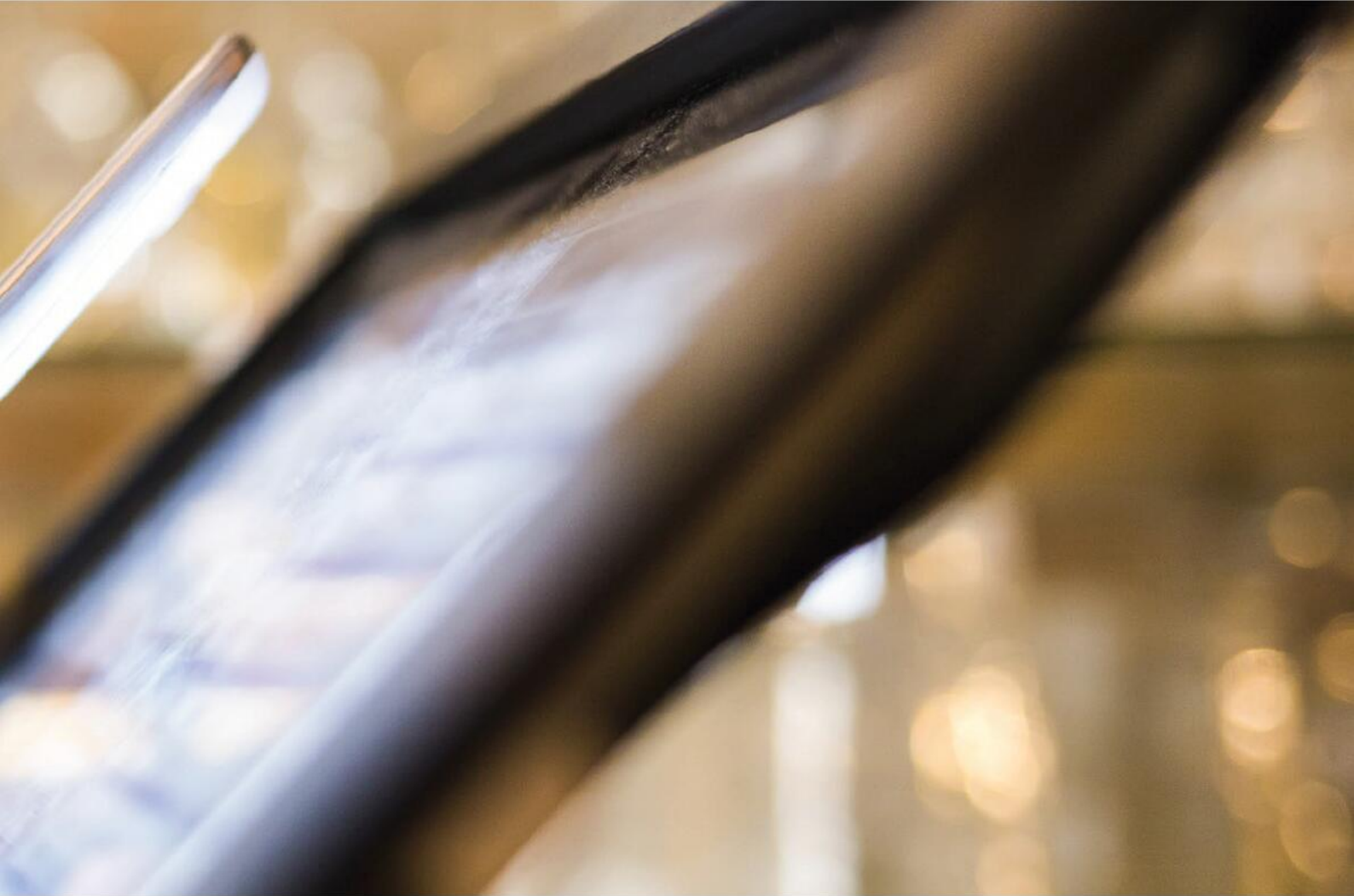
But how do you build and maintain trust in Artificial Intelligence?



## What is AI?

In our broad definition, AI is a collective term for computer systems that can sense their environments, think, learn and take action in response to what they're sensing and their objectives.

	Human in the loop	No human in the loop
Hardwired/specific systems	<p><b>Assisted Intelligence</b></p> <p>Using data and analytics to drive business insights within existing decisions and actions.</p>	<p><b>Automation</b></p> <p>Automating business processes previously performed by humans.</p>
Adaptive systems	<p><b>Augmented Intelligence</b></p> <p>New ways for computers and humans to collaborate in making better decisions and taking more effective actions.</p>	<p><b>Autonomous Intelligence</b></p> <p>Systems that are adaptive and can autonomously carry out tasks without human intervention.</p>



## AI adoption challenges

AI brings potentially huge opportunities to reduce costs, improve decision making and gain better insight into customer behaviour. But there's very little awareness of some of the challenges and risks associated with AI. We believe there are five key areas:



### Too much variety

AI has opened up a variety of techniques that need to be tailored to different types of problems. Organisations face challenges around identifying the right techniques, tools and talent to focus their efforts.



### Implementation approach

Companies will need to develop lean approaches to implementing AI by filtering through large amounts of data, piloting AI efforts before scaling and developing a test and learn culture.



### People expect magic

AI is often thought of as automatic 'magical' learning without human intervention, when it often requires a laborious process of acquiring and cleansing data, labelling, training and guiding the algorithm.



### Lack of transparency and explainability

AI models are often perceived as 'black boxes' due to their complexity, so their decisions can be difficult to understand or accept. It is too easy to pass bias from data availability and our own biases to AI – or simply because the process underlying the data itself is a product of inequality or other biases.



### Validity of use case

Use of AI needs to start with a business problem and effectively be supported with deep domain expertise to create a solution offering, rather than acquiring a tool and tailoring the problem to suit the tool.

## How we can help

### 1 Responsible AI

- We recognise the value of assurance in unleashing the potential of AI. We've developed our Responsible AI framework and assessment tool to strengthen clients' confidence in how to effectively deploy AI solutions and have trust in their outputs.
- Our Responsible AI approach and maturity model provides a practical mechanism for bringing these priorities together and ensuring effective monitoring and stewardship of AI outcomes. We believe that these foundations will enable clients to accelerate innovation and realise their AI vision.

### 2 Explainable AI

- The effectiveness of AI powered systems is limited by a machine's inability to explain thoughts and actions to users. Explainable AI is essential if users are to understand, trust, and effectively manage AI systems.
- Our Explainable AI proposition helps clients address the crucial task of demonstrating transparency across the AI ecosystem. This will include the ability to demonstrate regulatory compliance amongst other things.

### 3 AI risk assessment

- Supporting clients in finding the right balance between capitalising on opportunities that AI use cases can deliver and providing the appropriate level of risk management.
- Our AI evaluation framework helps quantify the different levels of risk associated with any AI use cases and, based on a client's risk thresholds, identify and implement the appropriate lines of defence, ranging from proactive monitoring through to explainable AI.

### 4 AI portfolio risk management

- As our clients increasingly adopt AI solutions across their value chain, they will need to adopt a more holistic approach for managing the use of AI across their enterprise.
- Our AI Portfolio risk management tool enables our clients to look at the how and where AI is being adopted and monitor the level of risk they are carrying across their core business processes as well as at an enterprise level.

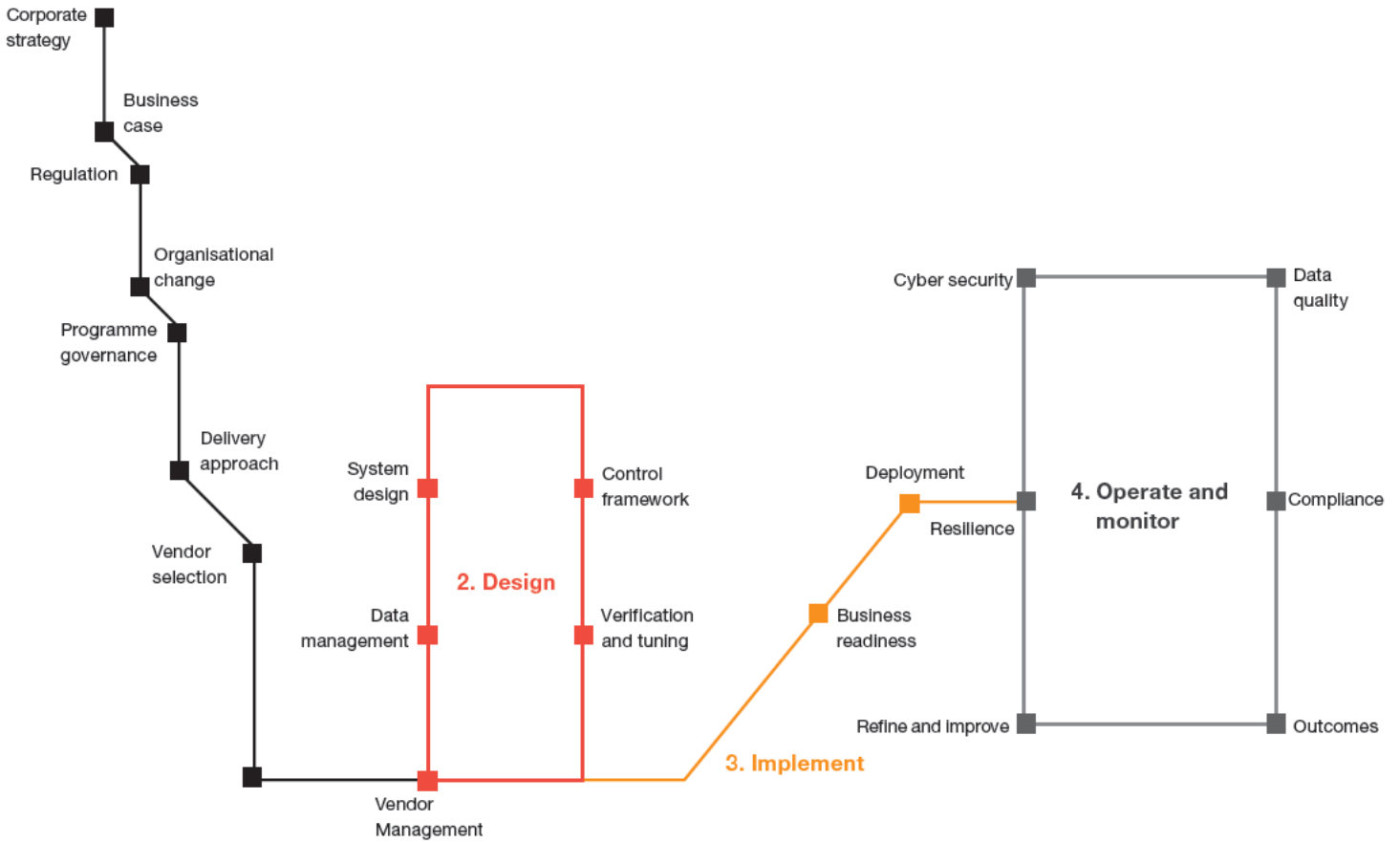
### 5 AI incident response

- We expect the number of AI related incidents to increase significantly as AI adoption increases significantly, resulting in both trust issues and reputational damage.
- Our AI Incident response service provides a rapid incident response diagnostic service to identify and explain the root causes for the incident and recommend the appropriate remediation activities.

# Responsible AI

Our **Responsible AI Framework** is designed to provide transparency and assurance over the responsibilities and explainability of AI development. It gives you and your stakeholders confidence that the necessary challenges and considerations are well thought through and that governance is in place to ensure that the business outcomes meet expectations.

## 1. Strategise

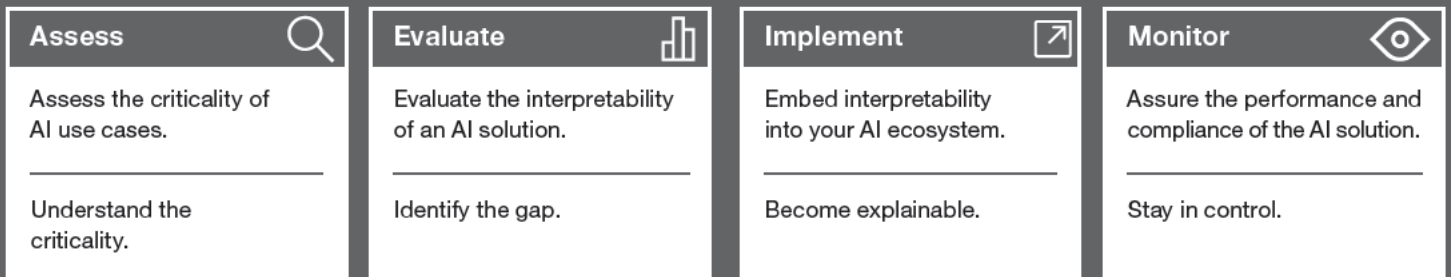


## Explainable AI

AI is growing in sophistication, complexity and autonomy, opening up transformational opportunities for business and society. At the same time, it makes explainability ever more critical.

As AI becomes more sophisticated, more and more decision making is being performed by an algorithmic 'black box'. To have confidence in the outcomes, build stakeholder trust and ultimately capitalise on the opportunities, it may be necessary to know the rationale of how the algorithm arrived at its recommendation or decision – 'Explainable AI'.

## How we can help



### Key outcomes

1. Provide stakeholder assurances over the use of AI
2. Be confident over the continued performance of the AI solution
3. Be able to demonstrate compliance

## Case studies

### Client



**Global financial services firm**



**UK central government department**

#### Context and background

A global financial services firm approached PwC to obtain assurances over their AI enabled loan approval system, prior to go-live. Their aim was to protect against risks of model bias which may result from correlations in data that discriminate against poorer classes of society or people with certain ethnicities. The organisation wanted to know if their AI solution was 'responsible' and fit-for-purpose.

One of the largest UK central government departments was exploring how AI could be used to improve the quality, efficiency and experience of how its public services were delivered to citizens. Following a series of proof of concepts the Department was struggling to identify transformational use cases and scale up their delivery capability to meet demands of the organisation.

#### What did we do?

Our team used our Responsible AI framework to perform a deep-dive assessment of the design, testing and implementation of the model. The assessment identified bias in the data that trained the AI model, resulting in age being the most influential factor in determining whether to approve a loan.

Our team used our Responsible AI framework to conduct a comprehensive review of the operating model, governance, and delivery approach. We developed a roadmap of operational improvement opportunities including a new, more robust governance model with greater business engagement, a focus on the strategic benefits to unlock funding and effectively manage demand across functions.

#### Business Impact

We highlighted the control weaknesses in the design and testing process which led to the bias issue, and provided recommendations for how to remediate and prevent from happening again. The bank is now aware of the design considerations for preventing bias in AI models and has established controls to ensure a fair and fit-for-purpose solution is developed.

The organisation revised its target operating model in order to support the department's growth strategy and current operations. Changes were made to governance mechanisms to assure that a responsible and controlled approach to the enterprise-wide adoption of AI. The wider stakeholder community increased their awareness of the potential benefits AI has to offer, and importantly where AI 'should' and 'should not' be used.

## Contacts



#### **Matthew White**

Partner, Digital Trust Leader

M: +971 (0)56 113 4205

E: [matthew.white@pwc.com](mailto:matthew.white@pwc.com)

[linkedin.com/in/mjwme](https://www.linkedin.com/in/mjwme)

@mjw0610



#### **Oliver Sykes**

Partner, Digital Trust

M: +971 (0)56 480 2447

E: [oliver.sykes@pwc.com](mailto:oliver.sykes@pwc.com)

[linkedin.com/in/osykes](https://www.linkedin.com/in/osykes)

@oraclesykes



#### **Clement Chan**

Director, ME AI Assurance Leader

M: +971 (0)50 152 3619

E: [clement.chan@pwc.com](mailto:clement.chan@pwc.com)

[linkedin.com/in/cclchan](https://www.linkedin.com/in/cclchan)

@Clem\_pwc