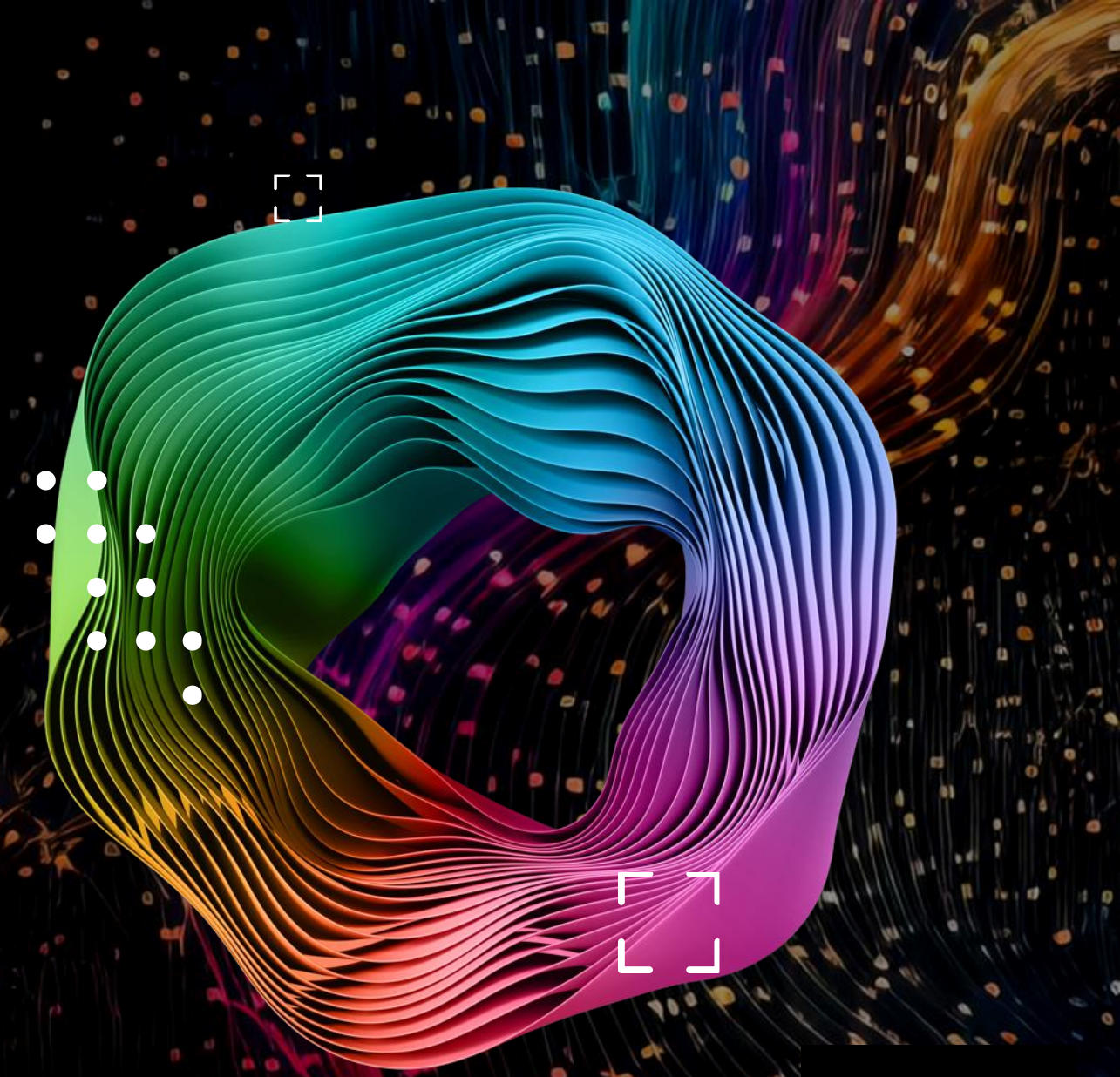
Deloitte.



A pragmatic approach to Trustworthy AI

Build and use AI in a Trustworthy way and comply with AI regulations

eu-LISA Industry Roundtable

EU Justice and Home Affairs in the Age of AI: Fostering Innovations and Mitigating Risks

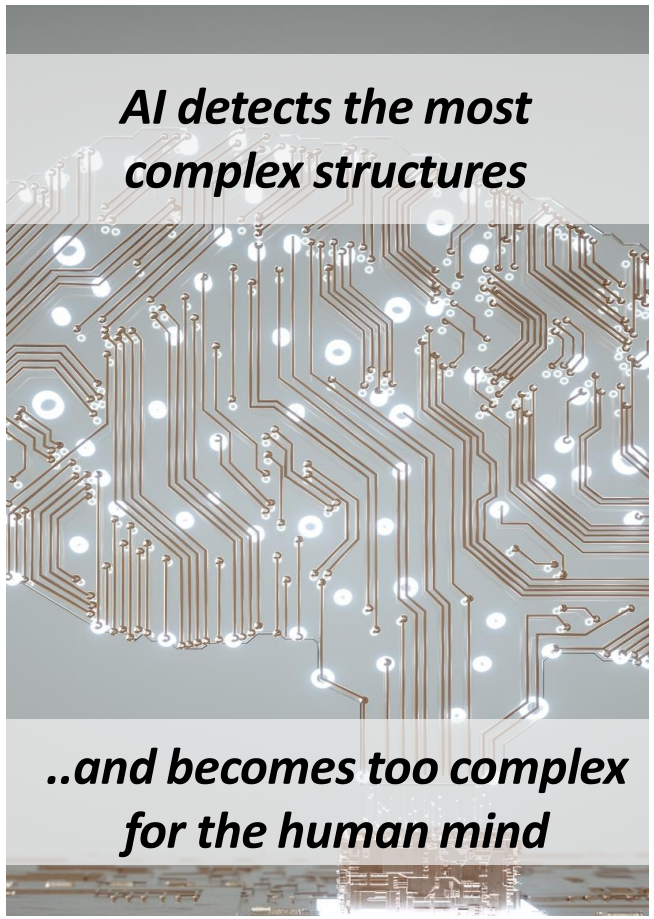
Budapest, November 12-13th, 2024

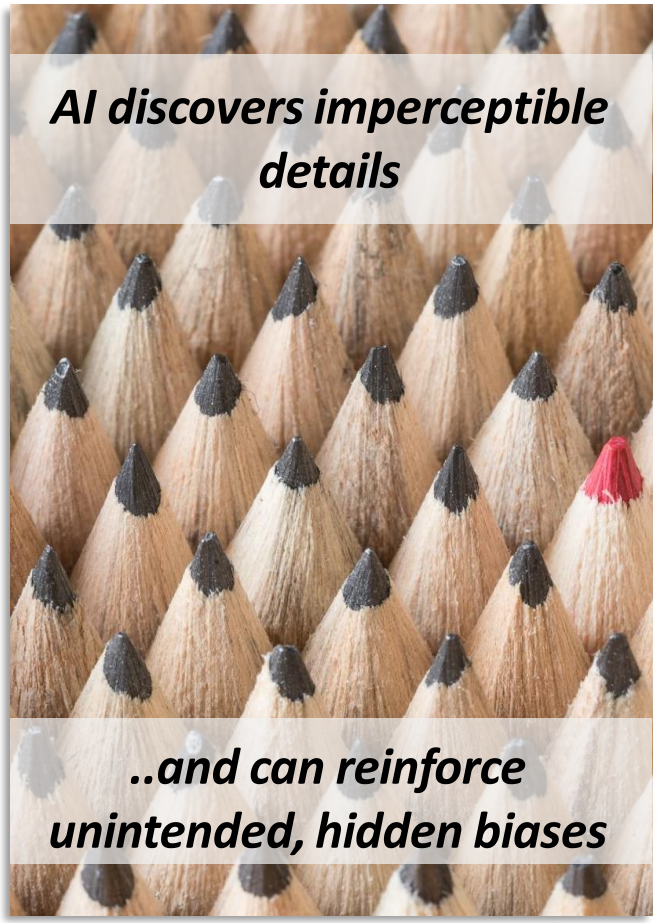
Lotte van den Berg & Hanne Verdickt, Trustworthy AI, Deloitte Belgium

Deloitte.

The Nature of AI comes with benefits and risks

Artificial Intelligence (AI) has a growing impact on our daily lives and can deliver exponential benefits to companies who can leverage its power effectively. But AI also comes with risks. In order to yield value, it requires people to trust its results. From the top-management to the end-users, everyone must be confident that AI is helping them.





Deloitte.

Addressing AI-related challenges with Trustworthy AI

Trustworthy AI is all about getting the confidence of people. Bringing trust in AI improves adoption and builds reputation – and is crucial to achieve the business impact.

Boost stakeholder trust in AI, increasing their rate of adoption of AI systems

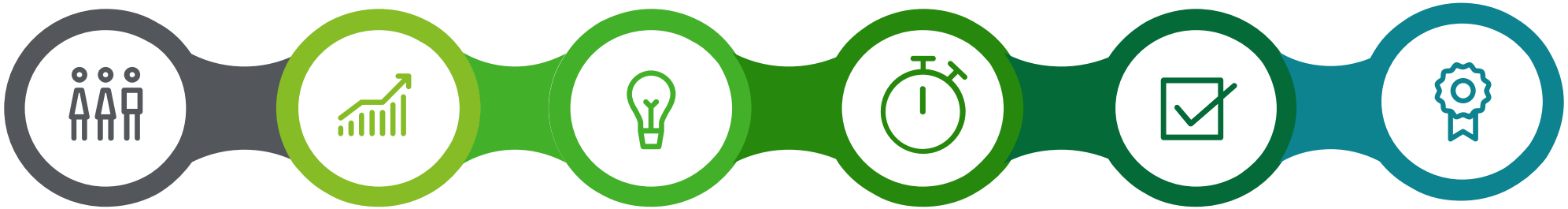
Elevate stakeholder adoption

Create transparent and streamlined processes that ease the way of working by creating structure and clarity

Accelerate your operations

Ensure you comply with the current regulatory framework, and are well positioned to adapt to future regulations

Meet compliance



Improve brand reputation and attractiveness

Elevate your brand's reputation at clients and attract talent with clear and open policies on AI

Enable innovation

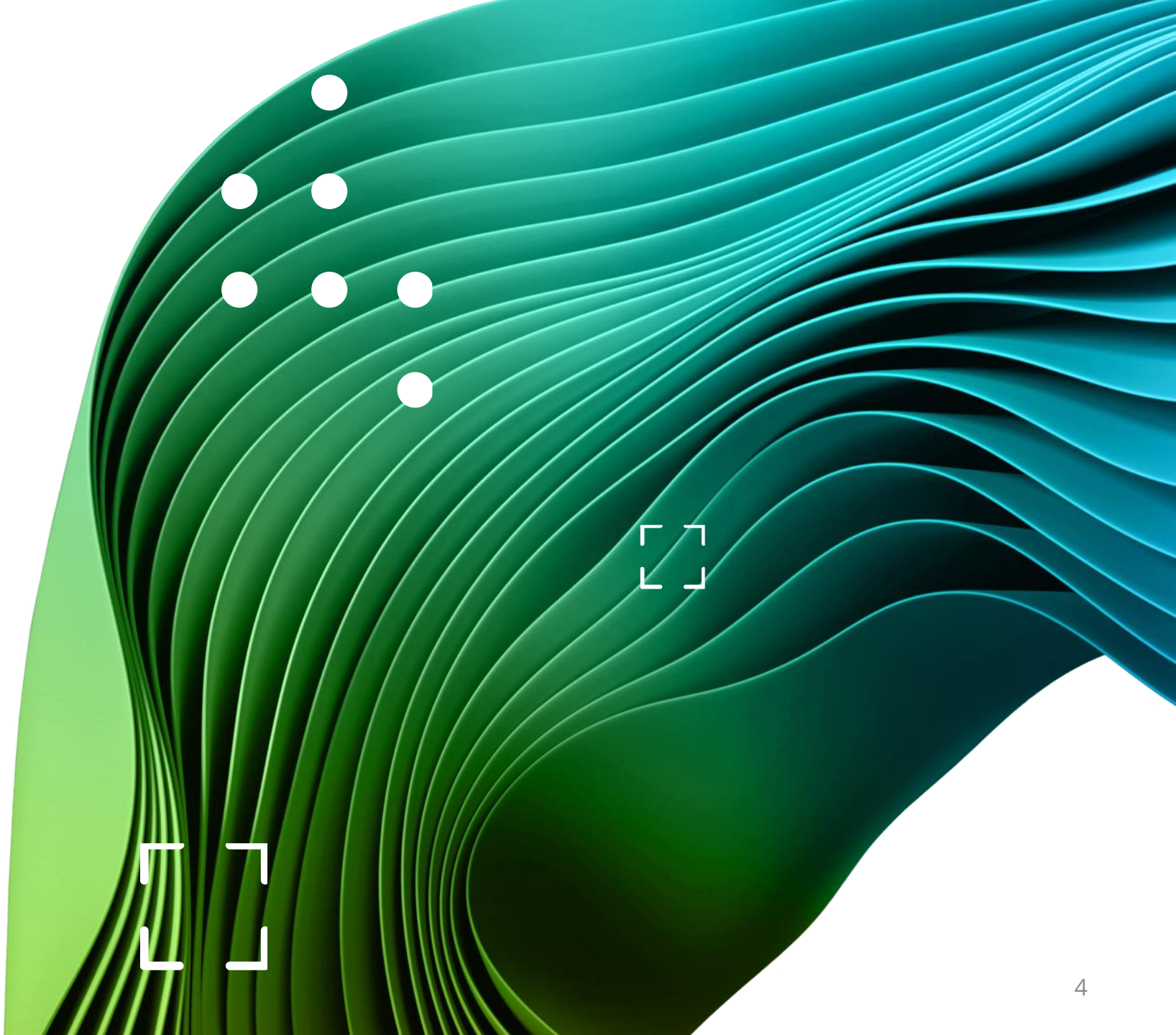
Allow employees to focus on innovation by striking the right balance between enforcing governance and creating an innovation playground

Enhance customer satisfaction

Improve customer outcomes by creating more transparent, consistent and understandable outcomes from your AI systems

To **unlock AI's great potential** and **enable the adoption of AI use cases**, organisations must ask …

"How can we leverage AI in a Trustworthy and Ethical way?"

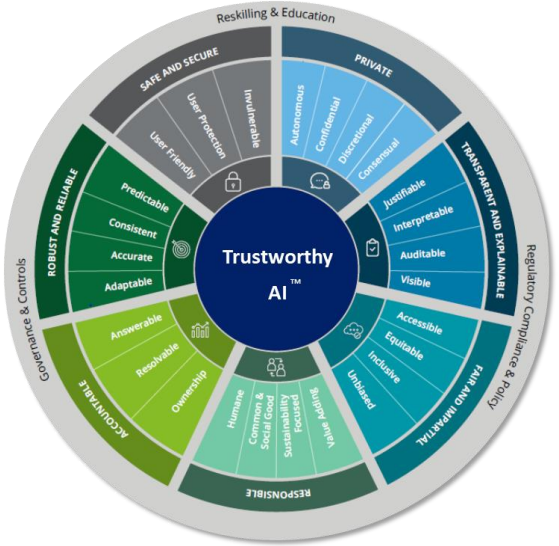


Deloitte.

What organisations need to build Trustworthy AI

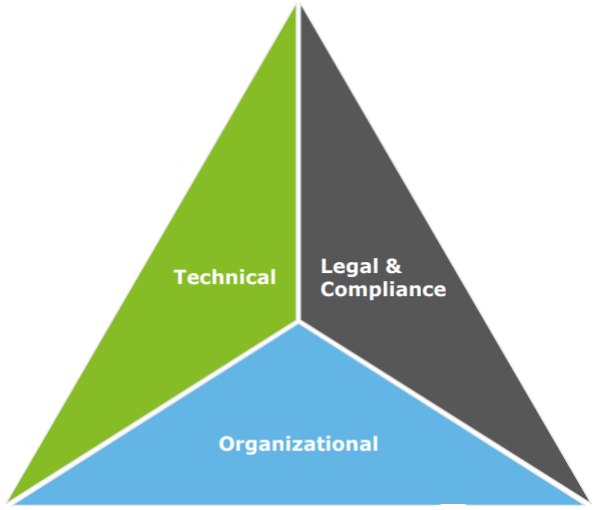
We follow a pragmatic and step-wise approach that is tailored to your organisation and balances the costs and benefits throughout the entire life cycle. By incorporating technical, organisational, risk, and compliance functions Deloitte helps building your Trustworthy AI Practice.

7 DIMENSIONS ON TRUSTWORTHY AI..



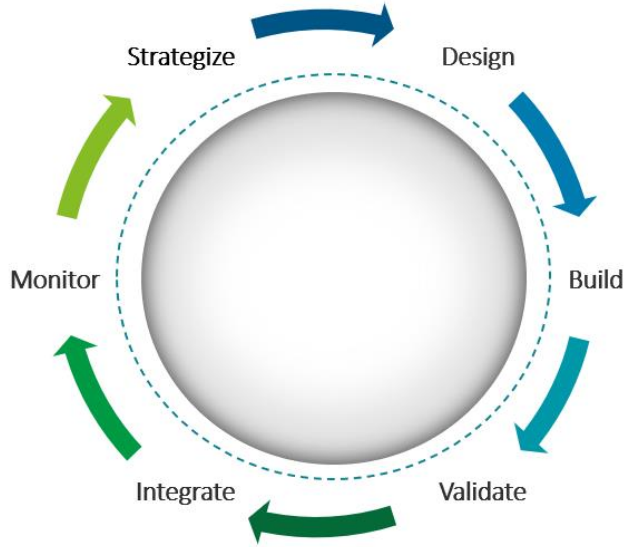
Encompassing Strategies & Objectives; Ideation & Requirements; Implementation & Enablement; Algorithm Assurance; Development & Systems; and Governance & Controls.

..ACROSS VARIOUS ORGANISATIONAL FUNCTIONS..



Deloitte's Trustworthy AI services extend beyond technical expertise to encompass communication, change management, advisory services, and other critical support elements.

..THROUGHOUT THE AI LIFE CYCLE



Implementing Trustworthy and Ethical AI in your organization throughout the entire AI lifecycle to enable AI solutions that are *Trustworthy By Design*.

Deloitte.

Guarding AI Trustworthiness **throughout the AI life cycle**

The proposed methodology is designed as a **gradual process**, in order to **support and improve the design, development, and adoption of AI systems in a pragmatic and efficient way**. The methodology questions guide the development team and serve as a checklist facilitating compliance with regulations, such as the AI Act. This approach enables to build out systematically the needed documentation of the AI System and develop Trustworthy AI solutions.

The process starts from the **scoping and design** phase where questions guide in the systematic evaluation of impacted stakeholders or business functions and an early risk identification as design requirement.

Development choices are explored and tested along the way, and solution design choices are documented along the way in the development phase.

Operate and ensure the performance and the model's reliability. **Monitor** business value, compliance and adoption.

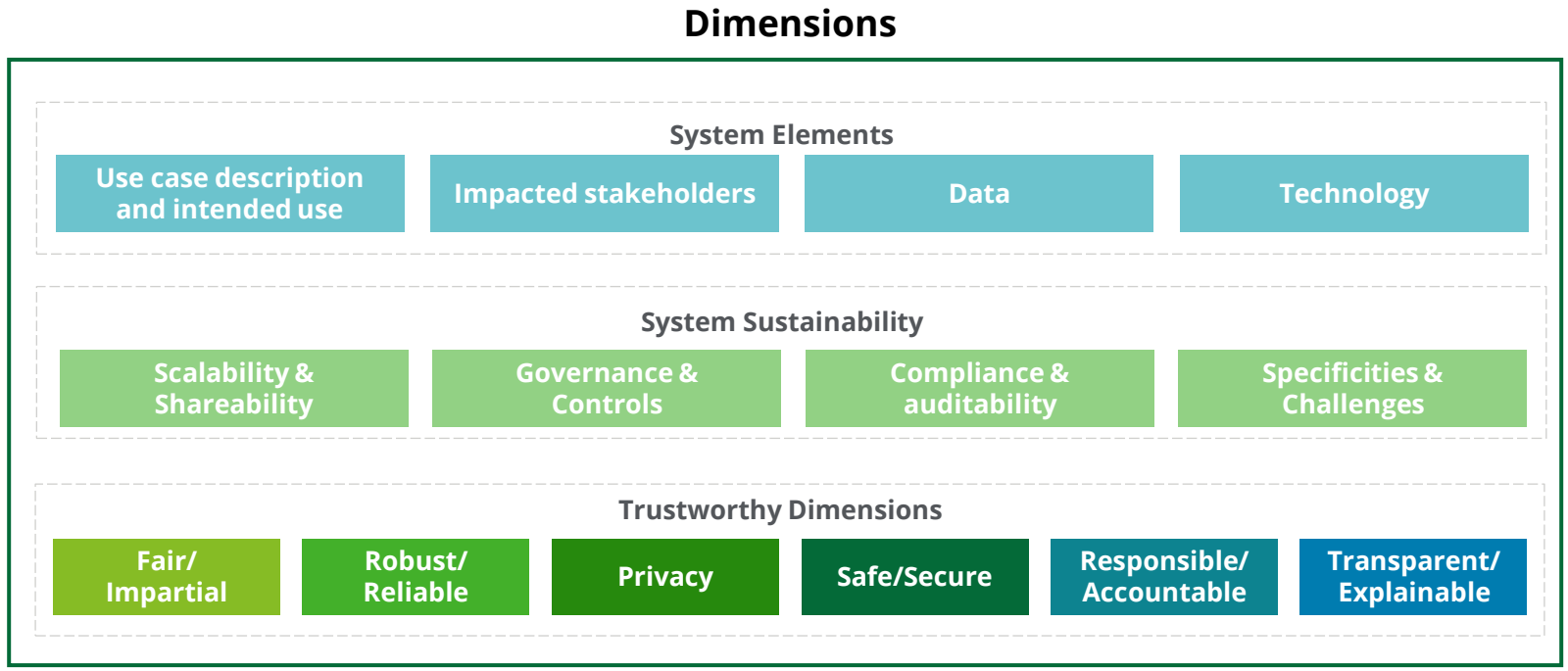
A critical stage in the process is the transition from the development phase to the operation phase, where the deployment of the AI solution is halted, unless all previous steps have been completed, and **the assessment output reaches acceptance**. The system can then **go live** and enter operations phase.

Deloitte.

Guarding AI Trustworthiness: **Assessment dimensions**

The AI assessment methodology consists of a set of questions that should be addressed to evaluate how the AI system performs in relation to crucial dimensions for the development of Trustworthy AI.

- The **AI assessment methodology** consists of a set of questions that should be addressed to evaluate how the AI system performs in relation to crucial dimensions for the development of Trustworthy AI (see figure on the right).
- The questions are structured according to the **AI lifecycle** (design, develop, operate) to assist the team in integrating Trustworthy AI components into their workflows.
- The result of the AI assessment is a **report** that presents an overview of the AI system's performance with respect to each dimension, emphasizing the areas with high risk levels that require attention.



Deloitte.

Dimensions – System Elements and Sustainability

There are 3 main dimensions to consider for AI governance: **AI trustworthiness, system sustainability and system elements**. These dimensions should always be considered along the AI lifecycle.

System Elements

Use case description and intended purpose	Identification of the problem to be solved or hypothesis to be investigated through a clear description of the use case and the path to undertake in achieving the intended purpose with a scoped solution.
Impacted stakeholders	From the use case description, define a list of impacted stakeholders who's interests and risks need to be taken into account throughout the system's lifecycle.
Data	Success of AI systems depends crucially on quality and availability of data, as well as the data type.
Technology	Transforming complexity into clarity through identifying the correct technology given the use case, stakeholders, and data available.

System Sustainability

Scalability & shareability	Ability to perform well under diversifying circumstances including increased workload or expanded use case contributes to the sustainability of the model.
Governance & controls	Monitor policies and guidelines, which describe roles and responsibilities, as well as the governance processes. AI practitioners are engaged appropriately and trained.
Compliance & auditability	Ensuring that the AI system is compliant with legislation such as the AI Act or Danish regulations and that it is controllable & auditable for authorities.
Specificities & challenges	Solution and context specific requirements or challenges that are relevant and need to be taken into account for a specific AI system, like requirements of e.g., domain-specific vertical legislations or related to a specific technology.

Deloitte.

Key Trustworthiness dimensions for AI

Applying Deloitte's 7-part Trustworthy AI Framework is an effective first step in diagnosing the ethical health of AI products while maintaining customer privacy, staying secure and abiding by relevant regulations.

SAFE & SECURE

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

ROBUST & RELIABLE

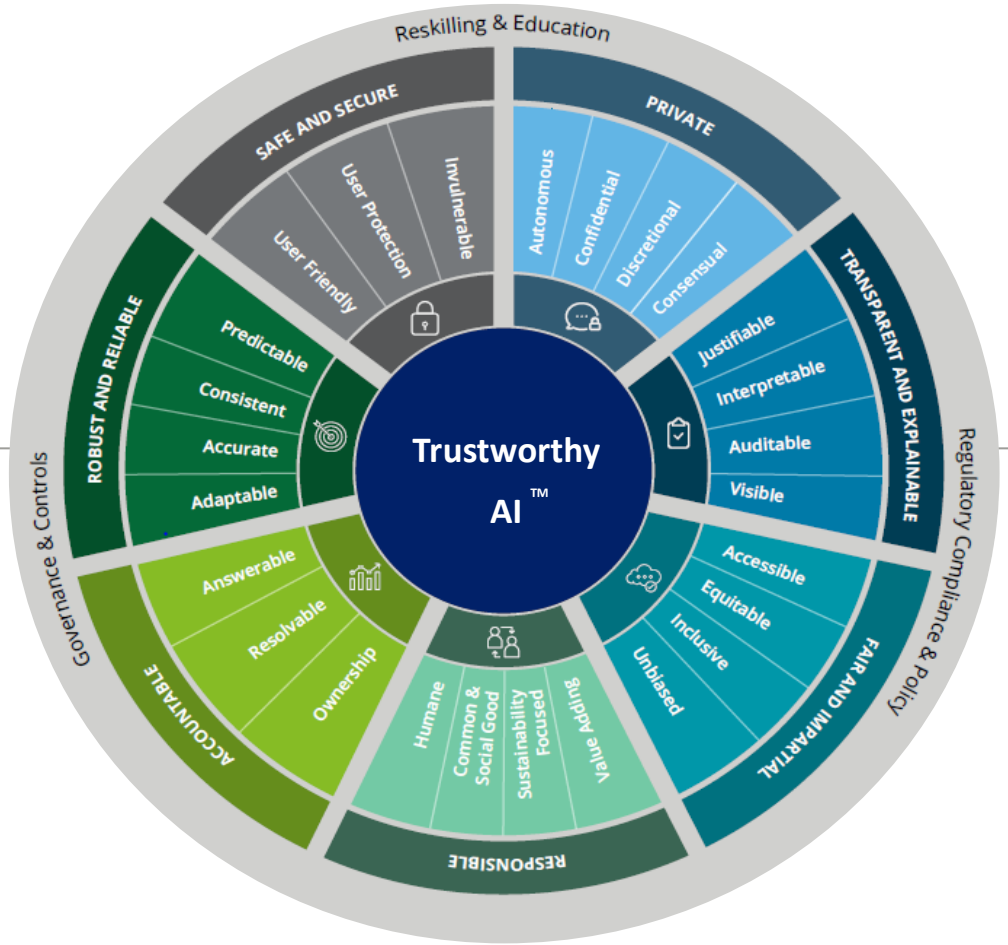
AI systems can learn from humans and other systems and produce consistent and reliable outputs

ACCOUNTABLE

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

RESPONSIBLE

The technology is created and operated in a socially responsible manner.



PRIVATE

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

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

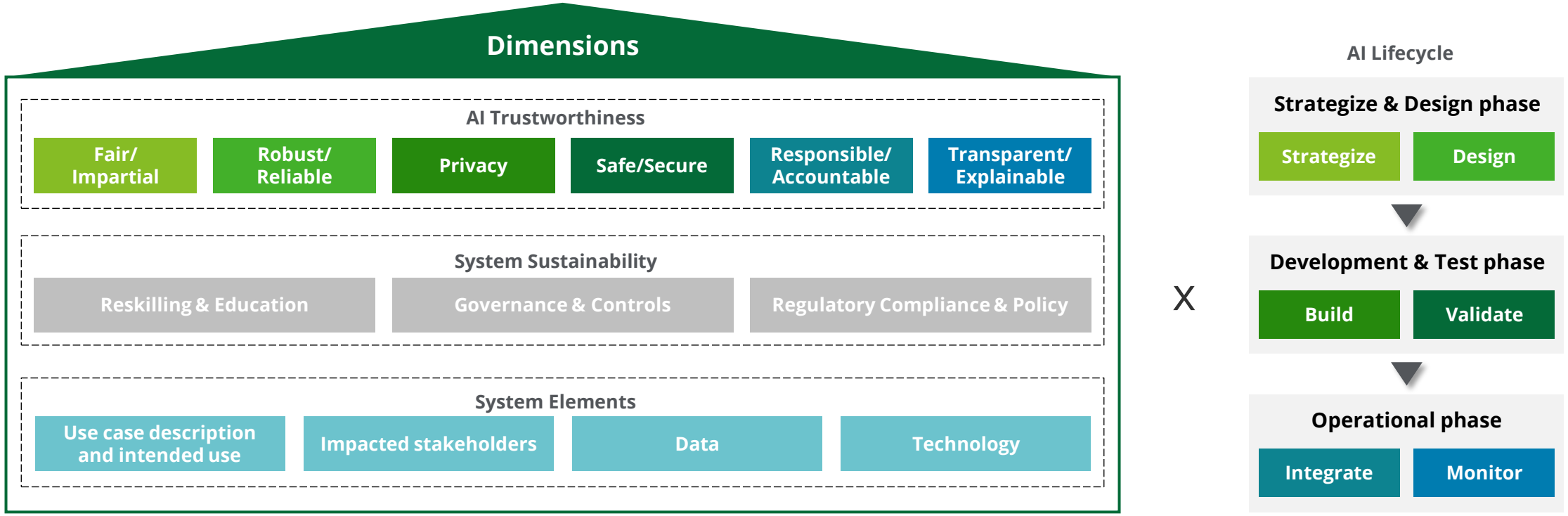
FAIR & IMPARTIAL

AI applications include internal and external checks to help ensure equitable application across all participants

Deloitte.

Deloitte Framework for AI Governance to achieve Trustworthy AI

There are 3 main dimensions to consider for AI governance: **AI trustworthiness, system sustainability and system elements**. These dimensions should always be considered along the AI lifecycle.



Deloitte.

Our proven expertise in supporting clients with Trustworthy AI

Leveraging our expertise in Trustworthy AI, legal, and assurance, we tailor robust AI solutions for your organisation, ensuring alignment with your specific needs. Our deep industry knowledge and cross-functional expertise uniquely enables us to design and implement effective Trustworthy AI strategies, drawing from insights and best practices of similar projects.

CONDUCT THOROUGH TRUSTWORTHY AI ASSESSMENTS

1 **Chemical company**
Conduct an AI Act Gap Assessment and elaborate a Trustworthy AI Roadmap

2 **Financial Services company**
Perform a risk assessment and ethical health check of their AI products

3 **EU DG Reform & Danish Gov**
Design a Trustworthy AI Assessment methodology and its governance

STRUCTURE THE COMPANY'S AI GOVERNANCE

4 **Public services company**
Define and operationalize AI Governance throughout the AI life cycle

5 **Higher Education company**
Deliver a strategy and a charter that sets the foundation for the Trustworthy use of data

6 **Consumer company**
Scope the AI Governance strategy and target operating model

DEPLOY TRUSTWORTHY AI PROJECTS

7 **Higher Education company**
Embed a Trustworthy AI approach to develop a use case on predicting student success

8 **Public services company**
Establish the AI Ethics Council to enable Trustworthy AI throughout the organisation

9 **Consumer company**
Provide detailed guidance on the AI Act, and other EU digital regulation, incl addressing interpretation queries for compliance assurance.

Deloitte.

Case 1: An AI Assessment methodology & for **Danish local, regional, and supervising authorities – DG REFORM**

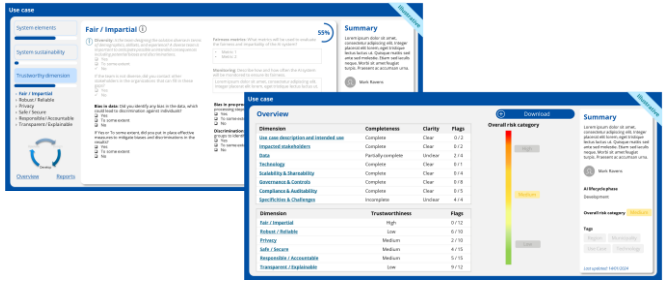
The Danish governments face the challenge of assessing the implications of adopting and operating AI systems from multiple perspectives (judicial, ethical, technical). To address this challenge, the project aims to **develop an AI Assessment Methodology adapted to the Danish context** to guide stakeholders in the oversight, development and deployment of Trustworthy AI.

Main deliverables of the project

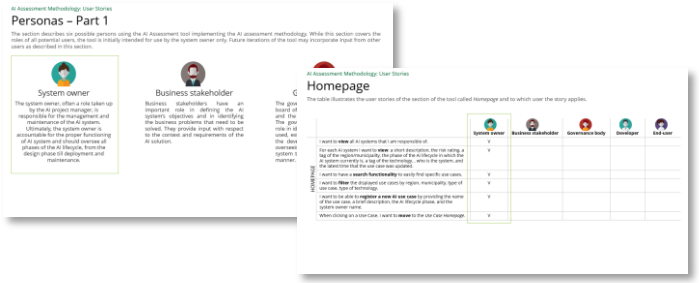
- ✓ AS-IS situation
- ✓ AI assessment benchmarking report
- ✓ Business case
- ✓ **AI Assessment Methodology** →
- ✓ TO-BE situation and SW architecture
- ✓ PoC
- ✓ Roadmap for implementation

AI Assessment Methodology

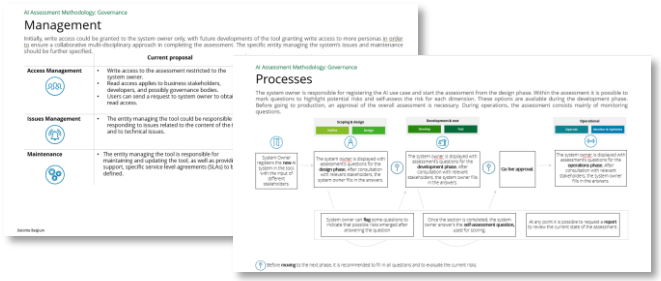
Structure



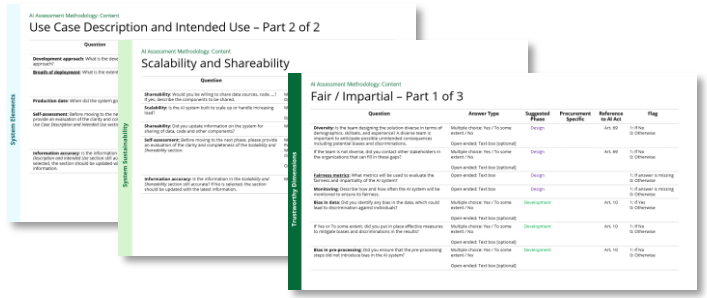
User Stories



Governance



Content



Deloitte.

Case 2: Trustworthy by Design for a **Belgian Public Employment Services organisation**

Design, tailor and operationalize the framework for a Public Employment Services organization for High-Risk AI Systems, aligned with the requirements from the AI Act

Preparations on the AI Act

The AI Act requirements were mapped, and a gap analysis was performed. Then recommendations were made to ensure compliance with the AI Act.

AI governance

In line with the gap analysis, the AI system lifecycle, the roles and responsibilities and the processes of AI governance were clarified. This was all delivered in a consolidated policy for the development and use of AI systems.
Furthermore, a risk management system was installed incl. a thorough methodology for risk identification

Documentation & pilot

All AI systems were thoroughly documented using a standard template to ensure transparency and explain-ability. Furthermore, an AI system risk registry was developed in which the AI risks are identified, described, evaluated and mitigated.

Ethics council

An Ethics Council was set up that provides advise on the ethical use of AI. Here, we focused on the operational model and setting up the governance process of the council. This included activities such as a stakeholder mapping and engagement plan, supporting the selection and engagement of internal and external Ethics Council members, agenda setting, kick-off meeting preparation.

Deloitte.

Case 3: AI governance strategy and roadmap for an international client active in chemicals & solutions

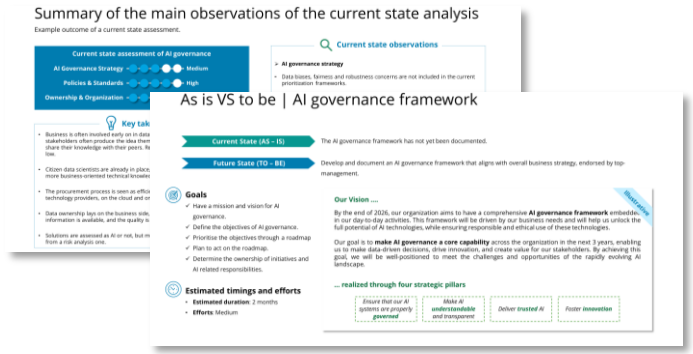
Helping a client active in chemicals and solutions with an assessment on their Trustworthy AI practices, including an AI Act gap analysis and building the AI Governance roadmap towards compliance.

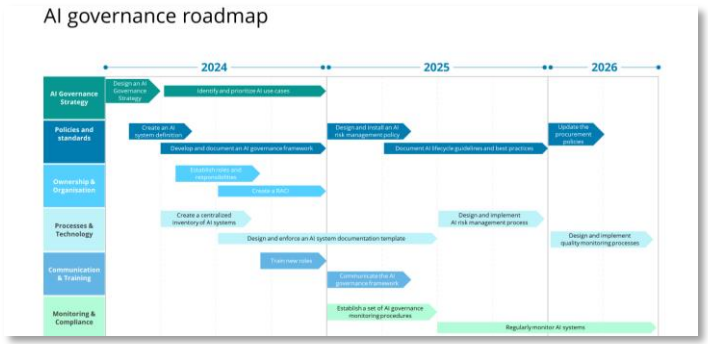
AI Governance & AI Act **gap analysis**

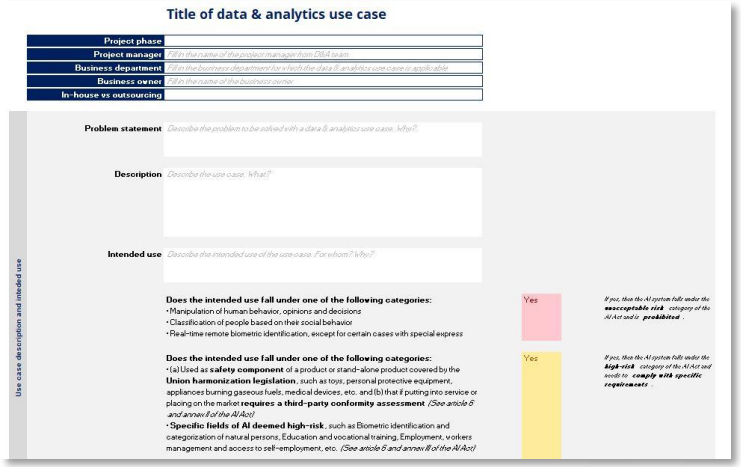
AI Governance **strategy**

AI Gov **Target Operating Model**

AI system **health check**







Thank you!

Lotte van den Berg
Trustworthy AI expert
✉ lovandenberg@deloitte.com

Hanne Verdickt
AI Governance expert
✉ hverdickt@deloitte.com