

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

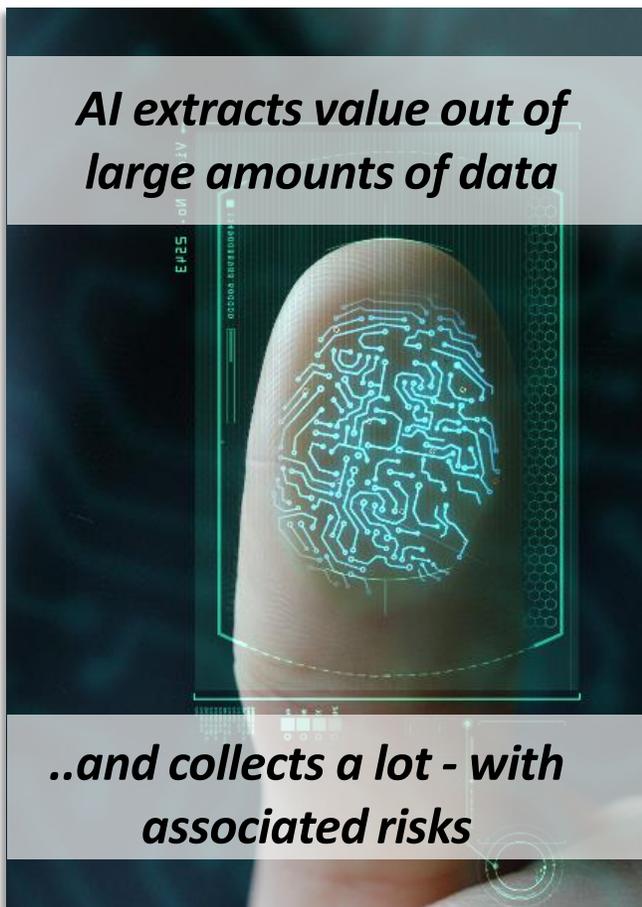


Trustworthy AI

The Nature of AI comes with benefits and risks

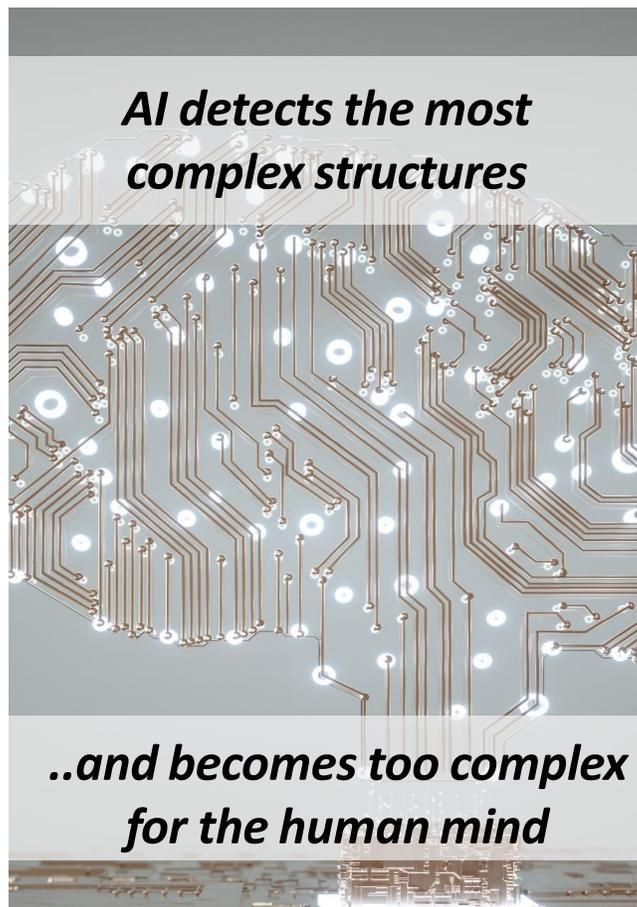
Artificial Intelligence (AI) has a growing impact on our daily lives and can deliver exponential benefits to companies who can leverage its power effectively. But AI also comes with risks. In order to yield value, it requires people to trust its results. From the top-management to the end-users, everyone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

AI extracts value out of large amounts of data



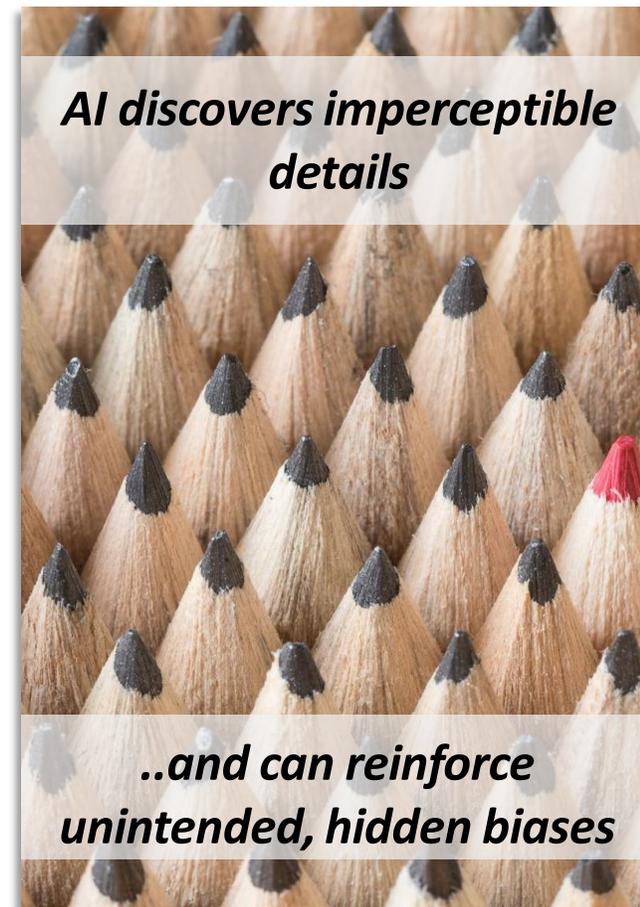
..and collects a lot - with associated risks

AI detects the most complex structures



..and becomes too complex for the human mind

AI discovers imperceptible details

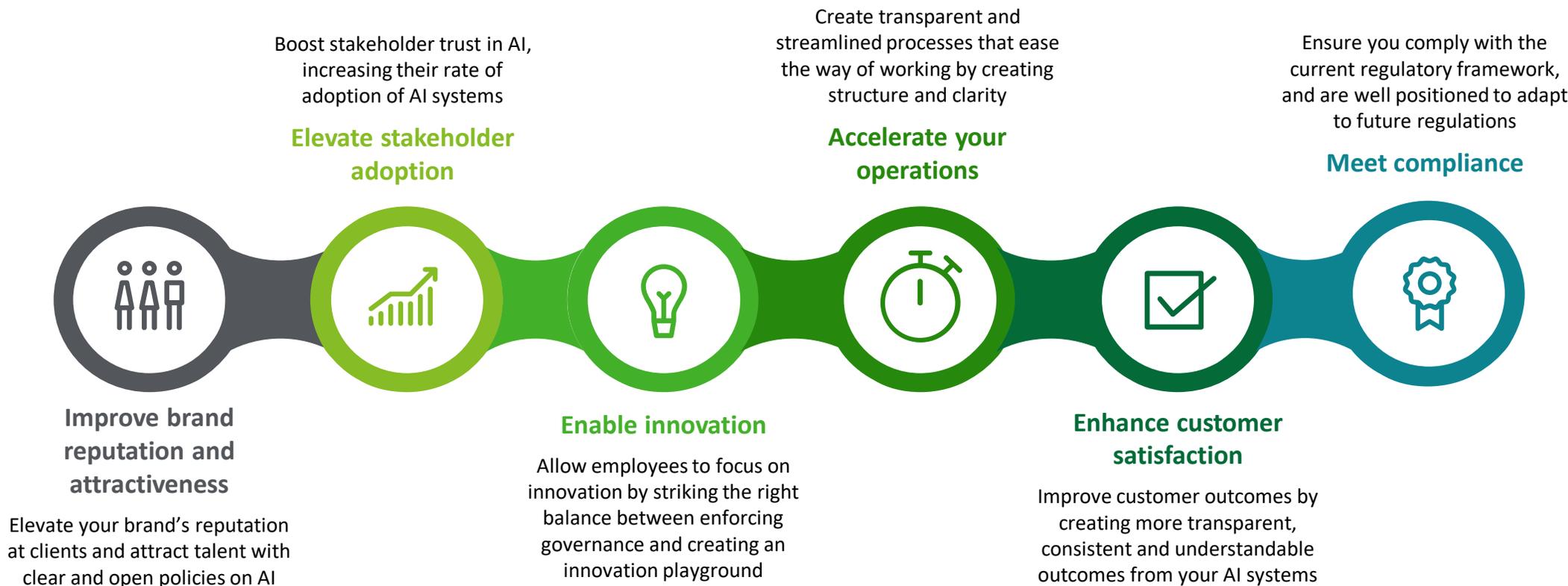


..and can reinforce unintended, hidden biases

Trustworthy AI

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.



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

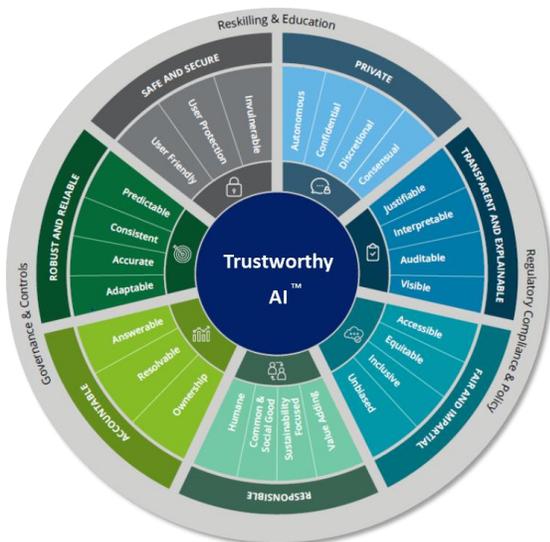
"How can we leverage AI in a Trustworthy and Ethical way?"

Trustworthy AI

What organisations need to build Trustworthy AI

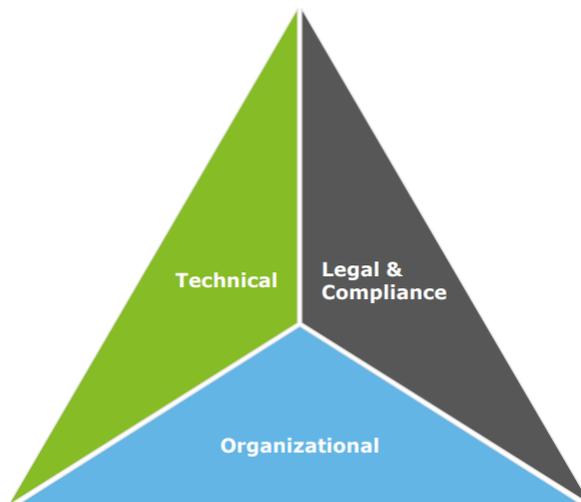
We follow a pragmatic and step-wise approach that is tailored to your organisation and balances the costs and benefits throughout the entire life cycle. By incorporating technical, organisational, risk, and compliance functions Deloitte helps building your Trustworthy AI Practice.

7 DIMENSIONS ON TRUSTWORTHY AI..



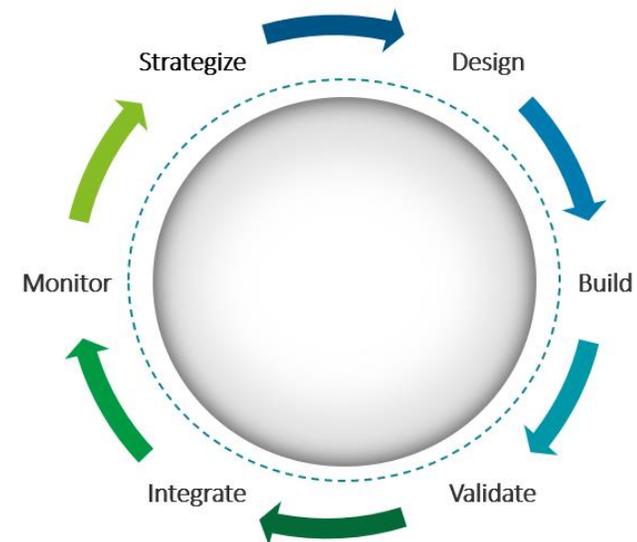
Encompassing Strategies & Objectives; Ideation & Requirements; Implementation & Enablement; Algorithm Assurance; Development & Systems; and Governance & Controls.

..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.

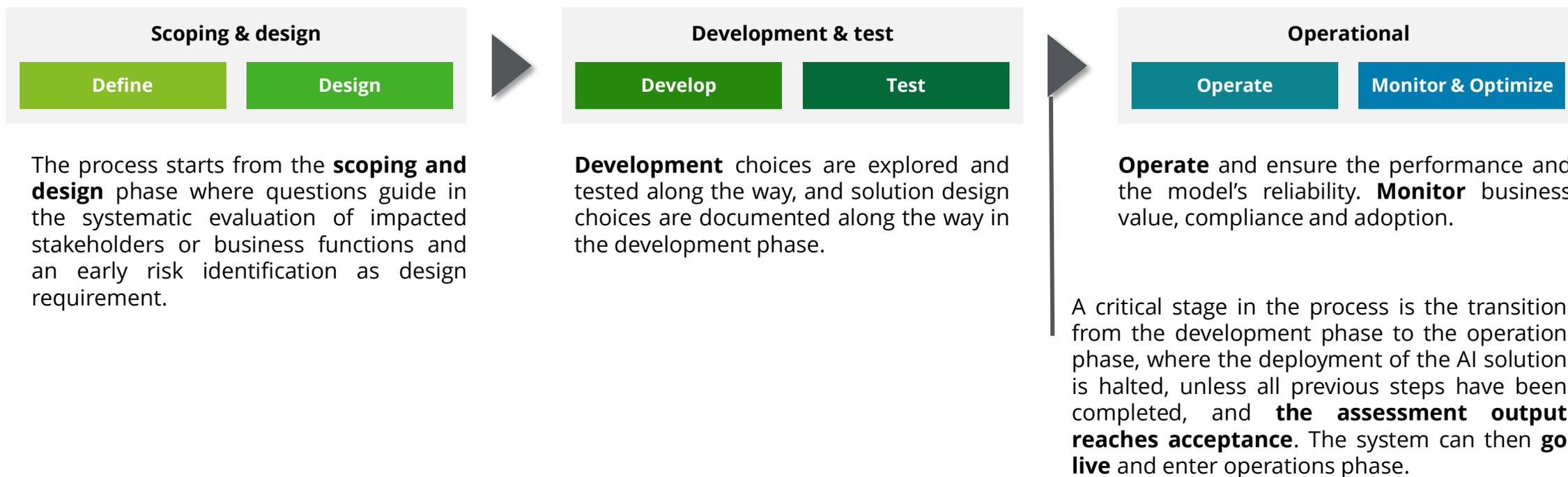
..THROUGHOUT THE AI LIFE CYCLE



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

Guarding AI Trustworthiness throughout the AI life cycle

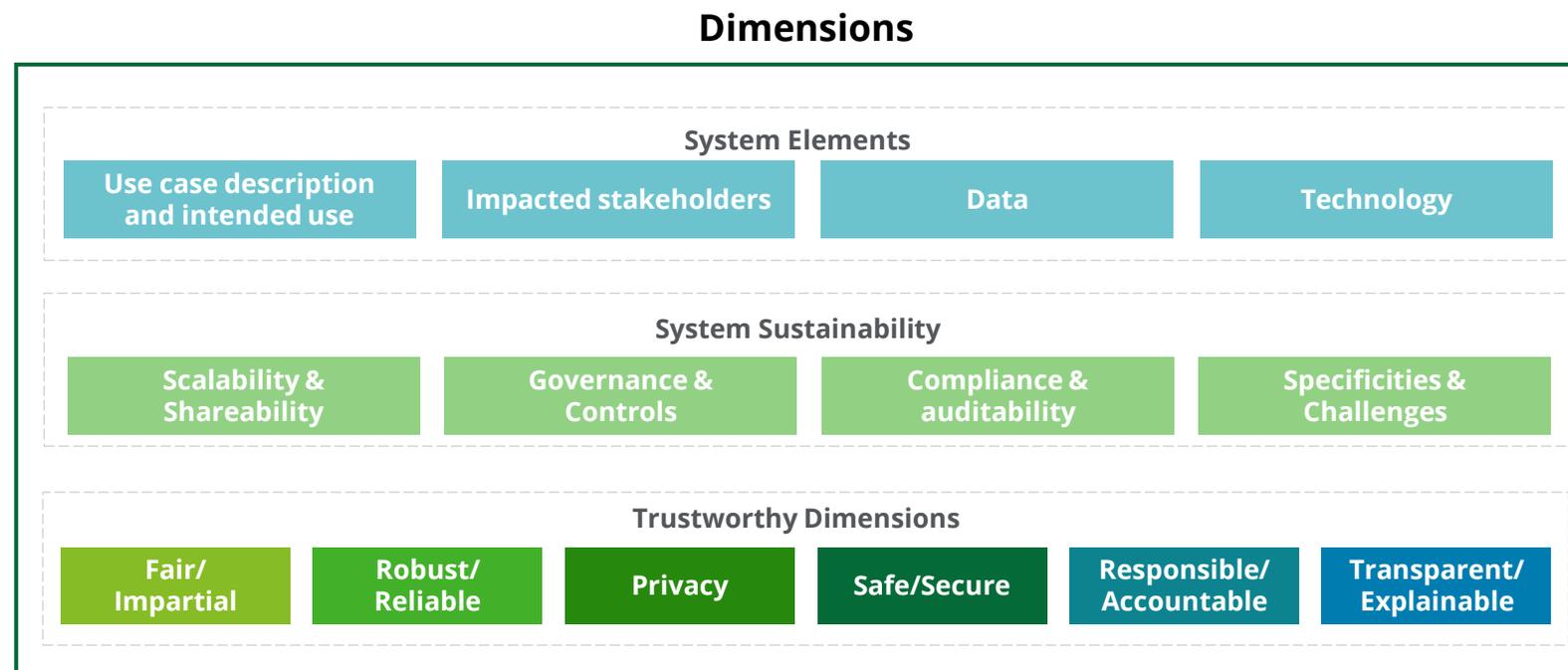
The proposed methodology is designed as a **gradual process**, in order to **support and improve the design, development, and adoption of AI systems in a pragmatic and efficient way**. The methodology questions guide the development team and serve as a checklist facilitating compliance with regulations,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



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.



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



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.

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.

Success of AI systems depends crucially on quality and availability of data, as well as the data type.

Transforming complexity into clarity through identifying the correct technology given the use case, stakeholders, and data available.

System Sustainability



Ability to perform well under diversifying circumstances including increased workload or expanded use case contributes to the sustainability of the model.

Monitor policies and guidelines, which describe roles and responsibilities, as well as the governance processes. AI practitioners are engaged appropriately and trained.

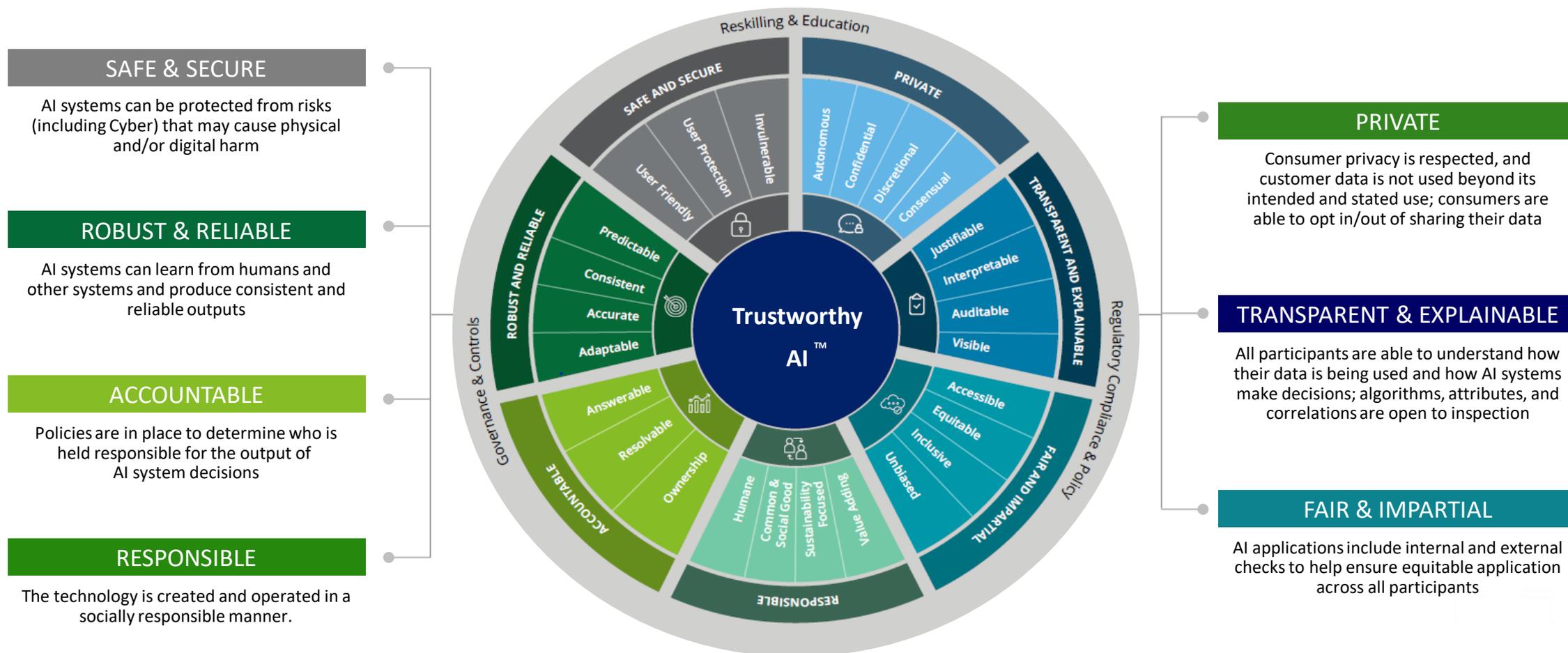
Ensuring that the AI system is compliant with legislation such as the AI Act or Danish regulations and that it is controllable & auditable for authorities.

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.

Trustworthy AI

Key Trustworthiness dimensions for AI

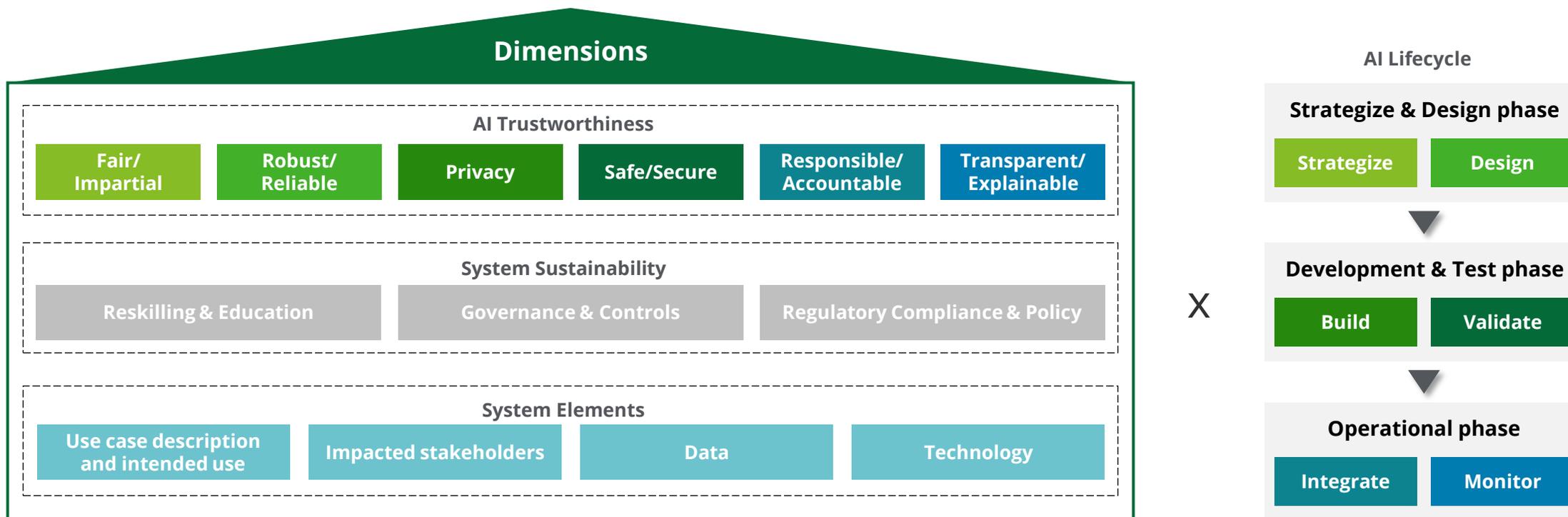
Applying Deloitte's 7-part Trustworthy AI Framework is an effective first step in diagnosing the ethical health of AI products while maintaining customer privacy, staying secure and abiding by relevant regulations.



Trustworthy AI

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.



Trustworthy AI in practice

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.

Trustworthy AI in practice

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

Dimension	Complexity	Clarity	Flags
Use case description and intended use	Complex	Clear	0/10
Goal	Complex	Unclear	2/10
Stakeholders	Complex	Clear	0/10
Benefits & Objectives	Complex	Clear	0/10
Context & Constraints	Complex	Clear	0/10
Complexity & Scalability	Complex	Clear	0/10
Specificities & Challenges	Complex	Unclear	4/10

User Stories

System owner
The system owner... [description]

Business stakeholder
Business stakeholder... [description]

Homepage
The table displays the user stories of the section of the tool called Homepage and to which user the story applies.

Governance

Management
Initial write access could be granted to the system owner only, with future developments of the tool granting write access to more personnel. It could be written a collaborative tool development approach to completing the assessment. The specific story managing the system issues and maintenance should be further specified.

Access Management
- Write access to the assessment restricted to the system owner.
- Read access applies to business stakeholders, developers, and possibly governance bodies...
- Users can send a request to system owner to obtain read access.

Processes
The system owner is responsible for registering the AI use case and start the assessment from the design phase. Within the assessment, it is possible to mark questions to highlight potential risks and self-assess the risk for each dimension. These options are available during the design phase before going to production, an approval of the several dimensions is required. During operations, the assessment includes reports of monitoring questions.

Content

Use Case Description and Intended Use – Part 2 of 2

Scalability and Shareability

Fair / Impartial – Part 1 of 3

Question	Answer Type	Suggested Format	Preparation Specifics	Reference	Flag
Self-assessment: Define a metric to measure the system's ability to scale and shareability.	Open-ended Text (Optional)	Design	None	641-89	0 - Data
Information assurance: Is the information assurance mechanism in place to ensure the system's ability to scale and shareability?	Open-ended Text (Optional)	Design	None	641-90	0 - Data
Information assurance: Is the information assurance mechanism in place to ensure the system's ability to scale and shareability?	Open-ended Text (Optional)	Design	None	641-91	0 - Data
Information assurance: Is the information assurance mechanism in place to ensure the system's ability to scale and shareability?	Open-ended Text (Optional)	Design	None	641-92	0 - Data
Information assurance: Is the information assurance mechanism in place to ensure the system's ability to scale and shareability?	Open-ended Text (Optional)	Design	None	641-93	0 - Data
Information assurance: Is the information assurance mechanism in place to ensure the system's ability to scale and shareability?	Open-ended Text (Optional)	Design	None	641-94	0 - Data
Information assurance: Is the information assurance mechanism in place to ensure the system's ability to scale and shareability?	Open-ended Text (Optional)	Design	None	641-95	0 - Data

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 explainability. 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.

Trustworthy AI in practice

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 Gov Target Operating Model



AI Governance strategy



AI system health check



Summary of the main observations of the current state analysis

Example outcome of a current state assessment.

Current state assessment of AI governance

AI Governance Strategy: Medium
Policies & Standards: High
Ownership & Organization: High

Key task

- Business is often involved early on in data stakeholders often produce the idea first then share their knowledge with their peers, to date.
- Open data scenarios are already in place, more business oriented technical business.
- The procurement process is seen as efficient technology providers, on the local and on.
- Data ownership lies on the business side, information is available, and the quality is
- Qualities are assessed as AI or not, but in future a risk analysis one.

Current state observations

AI governance strategy

- Data losses, biases, and obscures concerns are not included in the current procurement frameworks.

As is VS to be | AI governance framework

Current State (AS - IS) The AI governance framework has not yet been documented.

Future State (TO - BE) Develop and document an AI governance framework that aligns with overall business strategy, endorsed by top management.

Goals

- Have a mission and vision for AI governance.
- Define the objectives of AI governance.
- Prioritize the objectives through a roadmap.
- Plan to act on the roadmap.
- Determine the ownership of initiatives and AI related responsibilities.

Our Vision ...

By the end of 2026, our organization aims to have a comprehensive **AI governance framework** embedded in our day-to-day activities. This framework will be driven by our business needs and will help us unlock the full potential of AI technologies, while ensuring responsible and ethical use of these technologies.

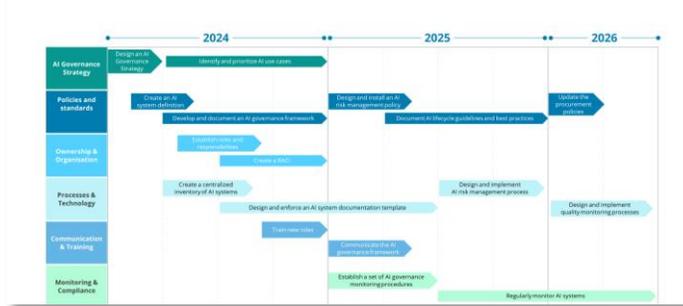
Our goal is to **make AI governance a core capability** across the organization in the next 3 years, enabling us to make data-driven decisions, drive innovation, and create value for our stakeholders. By achieving this goal, we will be well-positioned to meet the challenges and opportunities of the rapidly evolving AI landscape.

Estimated timings and efforts

- Estimated duration: 2 months
- Effort: Medium

Deliverables: Ensure that our AI systems are properly governed, Make AI understandable and transparent, Deliver trusted AI, Foster innovation.

AI governance roadmap



Title of data & analytics use case

Project phase	
Project manager	<i>(fill in the name of the project manager from ODA team)</i>
Business department	<i>(fill in the business department to which the data & analytics use case is applicable)</i>
Business owner	<i>(fill in the name of the business owner)</i>
In-house vs outsourcing	

Problem statement *(Describe the problem to be solved with a data & analytics use case. What?)*

Description *(Describe the use case. What?)*

Intended use *(Describe the intended use of the use case. For whom? Why?)*

Does the intended use fall under one of the following categories:

- Manipulation of human behavior, operators and decisions
- Classification of people based on their social behavior
- Real-time remote biometric identification, except for certain cases with special express

Does the intended use fall under one of the following categories:

- Used as **safety component** of a product or stand-alone product covered by the **Union harmonization legislation**, such as toys, personal protective equipment, appliances burning gaseous fuels, medical devices, etc. and that will be put into service or placed on the market requires a **third-party conformity assessment** (See article 6 and annex 8 of the AI Act)
- Specific fields of **AI deemed high-risk**, such as Biometric identification and categorization of natural persons, Education and vocational training, Employment, workers management and access to self-employment, etc. (See article 6 and annex II of the AI Act)

Yes **No**

*If yes, the AI system falls under the **unacceptable risk** category of the AI Act and is **prohibited**.*

*If yes, the AI system falls under the **high-risk** category of the AI Act and needs to **comply with specific requirements**.*

Use case description and intended use

Thank you!



Lotte van den Berg
Trustworthy AI expert

✉ lovandenberg@deloitte.com



Hanne Verdickt
AI Governance expert

✉ hverdickt@deloitte.com

