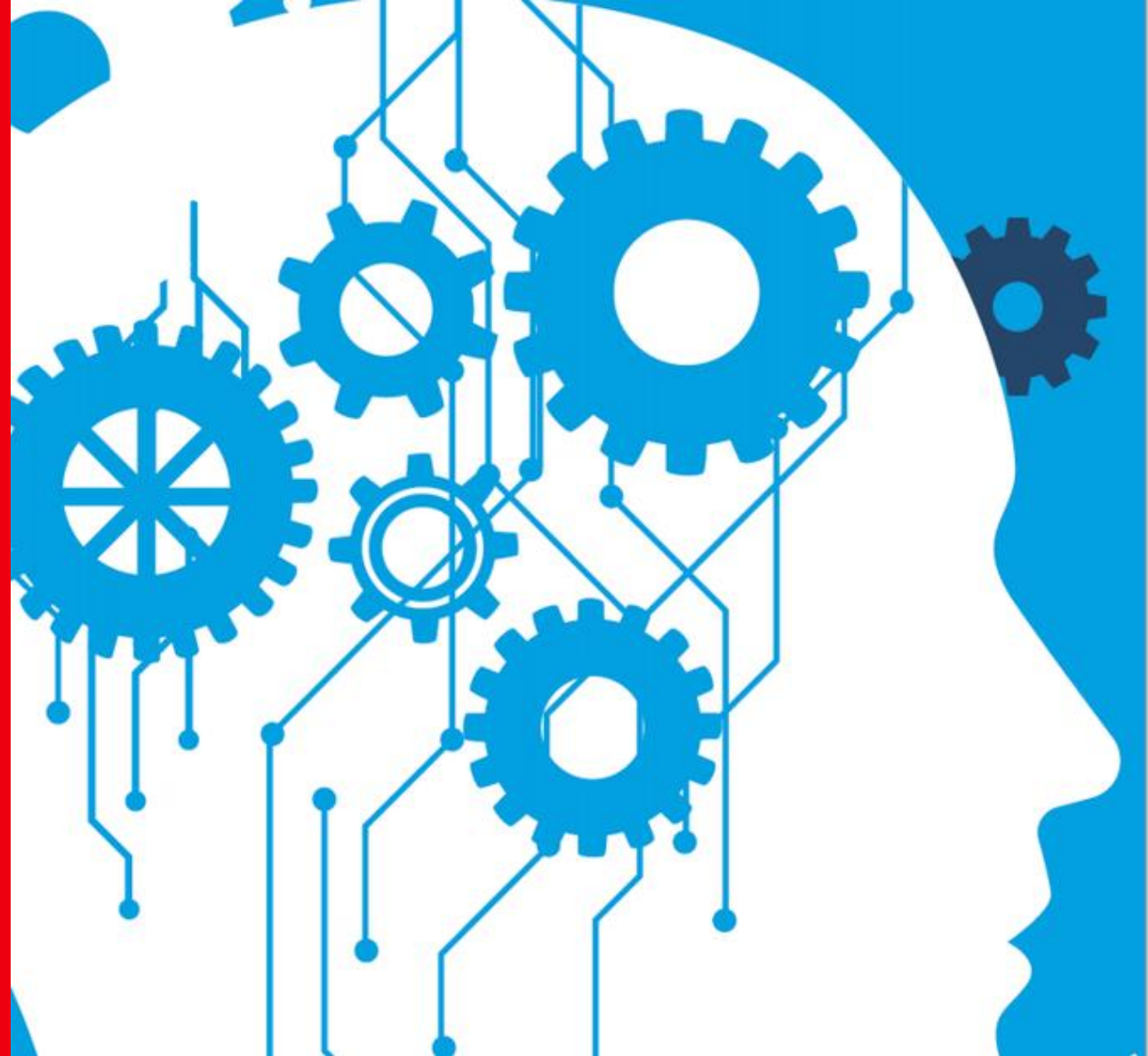# Explainable AI "by design"

Martin van den Berg
martin.m.vandenberg@hu.nl

Agenda

Introductie

Het onderzoek

Resultaten

Toepassing

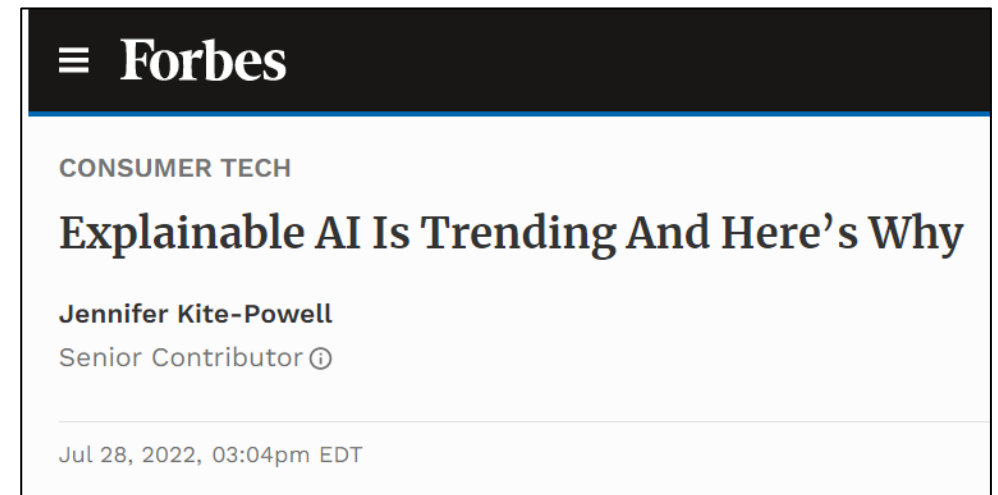Conclusie

What is explainable AI (XAI)?

Definition XAI:

"Given a stakeholder, XAI is a set of capabilities that produces an explanation (in the form of details, reasons, or underlying causes) to make the functioning and/or results of an AI system sufficiently clear so that it is understandable to that stakeholder and addresses the stakeholder's concerns."

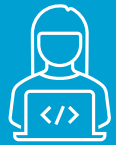https://www.hu.nl/onderzoek/publicaties/xai-in-the-financial-sector

Why XAI?

- Open up the black box

- Better understanding of model

- Manage model drift

- Improve models

- Manage risks of AI

- Increase trust and acceptance of AI

- Regulatory compliance (EU AI Act)

- Foundation of responsible AI

Forbes

CONSUMER TECH

Explainable AI Is Trending And Here's Why

Jennifer Kite-Powell

Senior Contributor ⓘ

Jul 28, 2022, 03:04pm EDT

https://www.forbes.com/sites/jenniferhicks/2022/07/28/explainable-ai-is--trending-and-h

"Which aspects play a role in the implementation of explainability of AI systems in the Dutch financial sector and how can these aspects be linked to the stages of the AI lifecycle?"

Category	Meaning	Level
Overall XAI	General policies, principles, and ways of working on XAI	Organization
Explainability and transparency in use case	Role and impact of explainability and transparency	Use case
AI in the use case	Role and impact of AI	Use case
Stakeholder's need for explanations in the use case	Stakeholders and their needs	Use case
XAI system in use case	Goal and approach of the XAI-system	Use case
Explanations in use case	What and how to explain	Use case
XAI methods and techniques in use case	Methods and techniques to develop the XAI system	Use case
Methods and techniques to evaluate XAI in the use case	Methods and techniques to evaluate the XAI system	Use case

Categories of aspects relevant in the design, development, and operation of XAI over the AI life cycle

Category	Meaning	Level
Overall XAI	General policies, principles, and ways of working on XAI	Organization
Explainability and transparency in use case	Role and impact of explainability and transparency	Use case
AI in the use case	Role and impact of AI	Use case
Stakeholder's need for explanations in the use case	Stakeholders and their needs	Use case
XAI system in use case	Goal and approach of the XAI-system	Use case
Explanations in use case	What and how to explain	Use case
XAI methods and techniques in use case	Methods and techniques to develop the XAI system	Use case
Methods and techniques to evaluate XAI in the use case	Methods and techniques to evaluate the XAI system	Use case

Categories of aspects relevant in the design, development, and operation of XAI over the AI life cycle

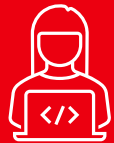| Stakeholder's needs for explanations in the use case	Check the stakeholders of the XAI system	Who are the stakeholder groups in need of an explanation (e.g., customers, regulators, internal officers, risk managers, senior management, model validators)?
	Check the stakeholder's needs for explanations	What are possible scenarios to prompt explanations (e.g., understanding inner workings, anticipating user questions, details about data, model mechanics at a high level, and ensuring ethical considerations during model development)?
		What are possible questions from stakeholders regarding explanations?
		What are the needs of stakeholder groups for explanations?

Category

Stakeholder's needs for explanations in the use case	Check the stakeholders of the XAI system	Who are the stakeholder groups in need of an explanation (e.g., customers, regulators, internal officers, risk managers, senior management, model validators)?
	Check the stakeholder's needs for explanations	What are possible scenarios to prompt explanations (e.g., understanding inner workings, anticipating user questions, details about data, model mechanics at a high level, and ensuring ethical considerations during model development)?
		What are possible questions from stakeholders regarding explanations?
		What are the needs of stakeholder groups for explanations?

Checkpoints

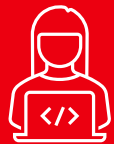| Stakeholder's needs for explanations in the use case	Check the stakeholders of the XAI system	Who are the stakeholder groups in need of an explanation (e.g., customers, regulators, internal officers, risk managers, senior management, model validators)?
	Check the stakeholder's needs for explanations	What are possible scenarios to prompt explanations (e.g., understanding inner workings, anticipating user questions, details about data, model mechanics at a high level, and ensuring ethical considerations during model development)?
		What are possible questions from stakeholders regarding explanations?
		What are the needs of stakeholder groups for explanations?

Questions and (possible answers)

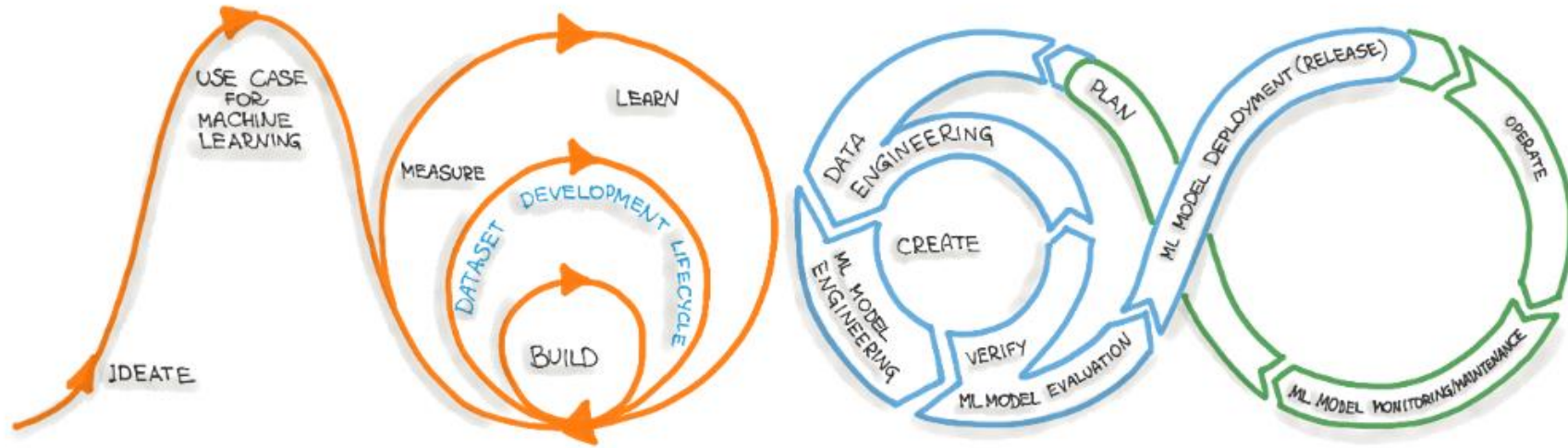| Stakeholder's needs for explanations in the use case	Check the stakeholders of the XAI system	Who are the stakeholder groups in need of an explanation (e.g., customers, regulators, internal officers, risk managers, senior management, model validators)?
	Check the stakeholder's needs for explanations	What are possible scenarios to prompt explanations (e.g., understanding inner workings, anticipating user questions, details about data, model mechanics at a high level, and ensuring ethical considerations during model development)?
		What are possible questions from stakeholders regarding explanations?
		What are the needs of stakeholder groups for explanations?

Category	Meaning	Level
Overall XAI	General policies, principles, and ways of working on XAI	Organization
Explainability and transparency in use case	Role and impact of explainability and transparency	Use case
AI in the use case	Role and impact of AI	Use case
Stakeholder's need for explanations in the use case	Stakeholders and their needs	Use case
XAI system in use case	Goal and approach of the XAI-system	Use case
Explanations in use case	What and how to explain	Use case
XAI methods and techniques in use case	Methods and techniques to develop the XAI system	Use case
Methods and techniques to evaluate XAI in the use case	Methods and techniques to evaluate the XAI system	Use case

Categories of aspects relevant in the design, development, and operation of XAI over the AI life cycle

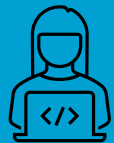| Explanations in use case	Check what to explain to whom	What are the contextual factors in providing explanations to stakeholders?
		What kind of information to provide as an explanation and to which stakeholders?
	Check how to deliver the explanation	How will the explanation be conveyed to stakeholders (e.g., in person, by a system)?
		What is the degree of interaction between the human and the machine in conveying the explanation (e.g., declarative, one-way interaction, two-way interaction)?
		What is the style of the explanation (e.g., text, visual)?
		What is the level of detail of the explanation (e.g., sparse, extensive)?
		What is the moment in time to provide the explanation (e.g., before or after the outcome)?
		How to give feedback if stakeholders inquire?

CRISP-ML(Q)

Checkpoint

Checkpoint

Checkpoint

Responsible:
Business Analyst
Accountable:
Product Owner

- Check what to explain to whom
- Check how to deliver the explanation

https://ml-ops.org/content/crisp-ml

In the form of Q&A

Question:
How should I use the checklist?

Answer:
There are different ways to make use of the checklist. The checklist can be used 'as is'. Second, the checklist can be integrated into the organization's AI lifecycle model and adjusted accordingly. In other words, the checklist can be embedded in the AI development process. This option is preferable when the organization already has a well-established AI development process. The advantage of this option is that the checklist can be tuned to the process and that the organizational roles can be aligned with the roles used in the organization or team. Next to that, the checkpoints and questions can be aligned with the terminology of the organization or team.
Furthermore, the checkpoints are helpful as a guide to document the decisions related to XAI.
Practical advice is to include the checkpoints as paragraph headings in the documentation template of the use case. Another practical piece of advice is to discuss beforehand which checkpoints are relevant to the use case. If in doubt, we recommend keeping the checkpoint relevant.

Question:
Do I always need to consider all the checkpoints and questions?

Answer:
The use of the checklist depends on the impact and risks of the use case. In low-impact or low-risk use cases, some checkpoints and questions may not apply. Our suggestion is to always go through all checkpoints and questions and determine whether they apply to a particular use case.

When explainability is required, many different aspects need consideration.

XAI is ~~...~~ AI development.

Explainable AI "by Design"

XAI must be integrated in the AI development process.

RAAK-mkb

In what ways can a meaningful explanation be generated and communicated to internal users of an AI system within financial services and how can it be evaluated whether that explanation meets the requirements of these users and applicable laws and regulations?

Thank you for your attention!