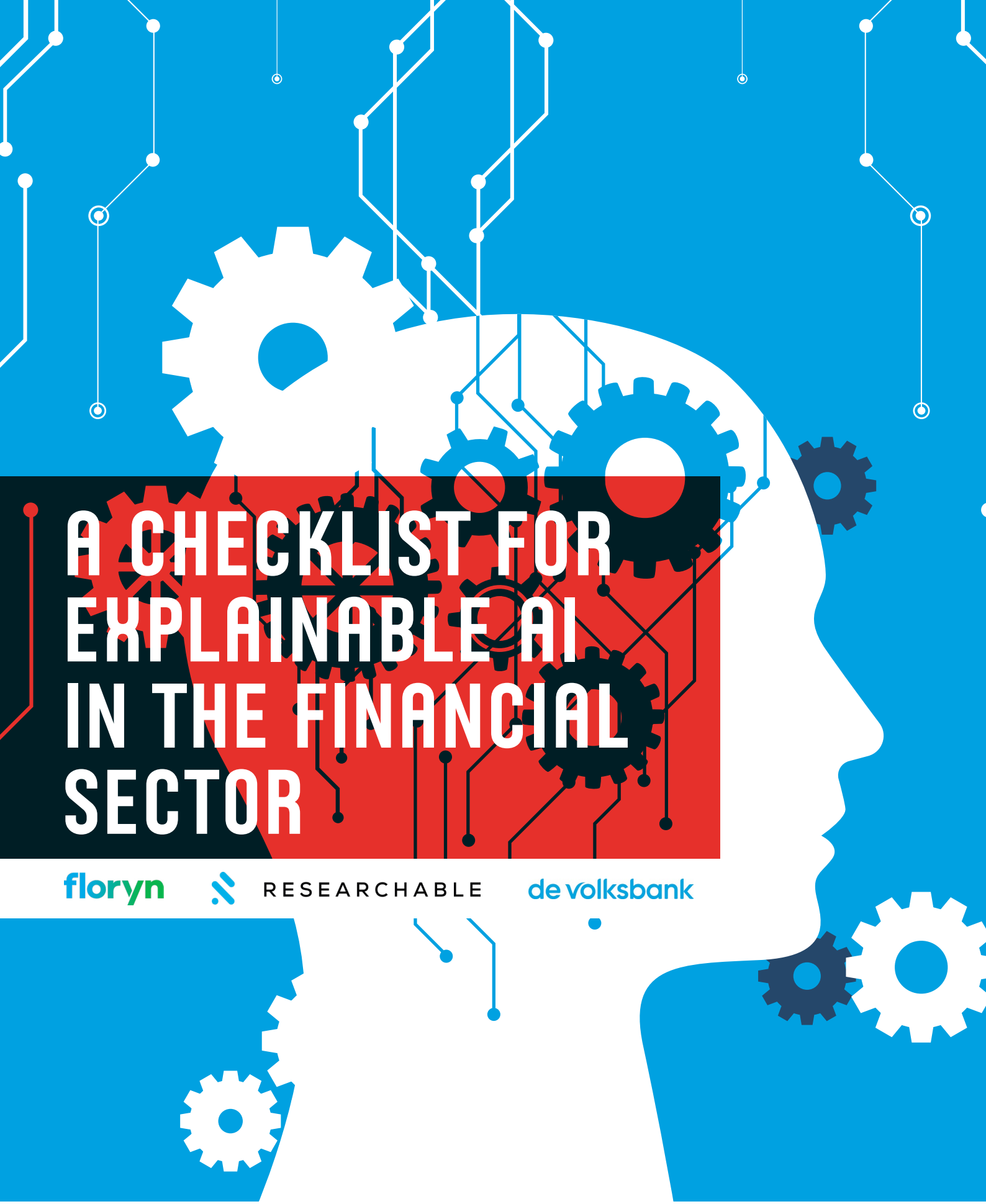


A CHECKLIST FOR EXPLAINABLE AI IN THE FINANCIAL SECTOR

floryn RESEARCHABLE de volksbank

HOGESCHOOL UTRECHT

IMPACT YOUR FUTURE

TABLE OF CONTENTS

PREFACE

This white paper is the result of a research project by Hogeschool Utrecht, Floryn, Researchable, and De Volksbank in the period November 2021-November 2022. The research project was a KIEM project[1] granted by the Taskforce for Applied Research SIA.

The goal of the research project was to identify the aspects that play a role in the implementation of the explainability of artificial intelligence (AI) systems in the Dutch financial sector. In this white paper, we present a checklist of the aspects that we derived from this research. The checklist contains checkpoints and related questions that need consideration to make explainability-related choices in different stages of the AI lifecycle. The goal of the checklist is to give designers and developers of AI systems a tool to ensure the AI system will give proper and meaningful explanations to each stakeholder.

We would like to thank our consortium partners Floryn, Researchable, and De Volksbank for their cooperation and willingness to share their experiences with us.

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

Yvette van der Haas, MSc
yvette.vanderhaas@hu.nl

Dr. Ouren Kuiper
ouren.kuiper@hu.nl

Dr. Marieke Peeters
marieke.peeters@hu.nl

Hogeschool Utrecht - Lectoraat Artificial Intelligence
November 2022 - Version 1.0

1 Case number KIEM.K21.01.046, date 24 November 2021.

1. INTRODUCTION

Artificial Intelligence (AI) is a key technology that will have a major impact on people and organisations (Scientific Council for Government Policy, 2021; Dutch Digital Delta, 2022). AI offers organisations unprecedented opportunities to operate more efficiently and effectively. However, there are risks associated with the use of AI. One of those risks is that the functioning of an AI system is so complex that it is no longer possible to explain how that AI system arrived at a certain prediction or decision. Explainable AI (XAI) focuses on generating explanations so that AI systems remain transparent and understandable (Adadi & Berrada, 2018; Arrieta et al. 2020). Explainability is also one of the ethical guidelines for trustworthy AI (HLEG, 2019). With the announced new EU AI Act from the European Commission (EC, 2021), explainability is expected to become a requirement for high-risk AI systems.

AI, and more specifically machine learning (ML), is being used increasingly in the Dutch financial sector. Both large financial institutions and fintechs use AI and ML in, amongst others, lending, customer acceptance, handling claims, and combating financial crime (Van der Burgt, 2019). McWaters (2019) notes that the opacity of AI systems poses a serious risk to the use of AI in the financial sector: lack of transparency can lead to a loss of control by financial institutions, thereby damaging consumer and societal trust. Given the crucial role of trust in the financial sector, explainability of the outcomes and functioning of AI systems is considered necessary (McWaters, 2019).

HU University of Applied Sciences Utrecht (HU) has been conducting practice-oriented research into XAI since 2020, focusing on the financial sector. In the first exploratory study in 2020, a framework was developed that was applied in a project of DNB's iForum with several banks (Kuiper et al., 2021; Van den Berg & Kuiper, 2020). This project showed that there is a need in the financial sector for a more detailed approach to implementing XAI. Scientific literature supports this need (Gerlings et al., 2020) whereby the successful implementation of XAI should address both technical (how to integrate explainability into an AI system) and social (how to integrate explanations into decision-making processes and how to communicate explanations with stakeholders) aspects (Bauer et al., 2021; Liao & Varshney, 2021; Kemper & Kolkman, 2019).

The demand for an approach for XAI prompted the HU to apply for funding for a research project (KIEM-regeling). The application was granted and starting from November 2021 a consortium of HU, Floryn, De Volksbank and Researchable investigated the following research question: 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?

This white paper contains the results of the research project and in particular, a checklist of aspects that we derived from our research. The checklist contains checkpoints and related questions that need consideration to make XAI-related choices in different stages of the AI lifecycle. The goal of the checklist is to give designers and developers of AI systems a tool to ensure the AI system will give proper and meaningful explanations to each stakeholder.

This white paper is structured as follows. In Section 2 we present related work as well as the definitions that we used in this research. Section 3 contains the research approach. In Section 4 we present the checklist and in Section 5 we discuss how to make use of the checklist. Finally, Section 6 contains the conclusions and a call to action.

2. RELATED WORK

AI is increasingly used in financial services. A sector research report from ING (2020) shows that AI has the most value for the IT sector and business and financial services. Both large financial institutions and SMEs use AI in, among other things, lending, customer acceptance, handling claims, and combating financial crime (Van der Burgt, 2019). With the increasing use of AI, there is also an increasing focus on XAI, which is seen as a means to ensure that AI systems remain transparent and understandable to stimulate the implementation and adoption of AI (Adadi & Berrada, 2018; Arrieta et al., 2020). This Section provides an overview of related work regarding the application of XAI in general, the application of XAI in the financial sector, and AI lifecycles.

2.1 Application of XAI in general

The explainability of AI systems is seen as one of the building blocks of the responsible use of AI (HLEG, 2019; Morley et al., 2019). The essence of responsible AI is about values such as respect for human autonomy, the prevention of harm and bias, non-discrimination, fairness, and explainability. These values are reflected in guidelines for the responsible use of AI such as European guidelines (HLEG, 2019). XAI is an essential foundation for the responsible use of AI because it helps explain how an AI system works and, as such, supports the detection of bias, fairness, and discrimination.

AI systems differ greatly in complexity. AI systems based on, for example, regression or decision trees, are easy to understand and the operation is relatively easy to explain. Even for a user unfamiliar with the technology, it is easy to see which variables play an important role in the outcome of the model. The operation of models based on such algorithms is relatively easy to explain. With the rise of new kinds of models, such as random forest and deep neural networks (DNNs), AI models are becoming increasingly complex. The reason these new kinds of models are used is that in many situations they outperform conventional models such as regression or decision trees. These new kinds of models are considered complex black-box AI models; models whose operation is not inherently apparent from the model itself. The opposite of a black-box model is a transparent model, i.e., a model for which its operation is easy to understand (Arrieta et al., 2020). As new kinds of AI models are increasingly used to make important predictions in high-impact use cases, the need for transparency has increased, and with it the need for XAI (Islam et al., 2022). One of the main goals of XAI is to provide a solution to the black-box problem in AI by making complex AI systems more transparent. But XAI goes further than just uncovering the internal aspects AI systems. Another goal of XAI is to increase the trust and adoption of AI systems by stakeholders such as customers, regulators, and users. This can be achieved by examining to what extent people understand the decisions of an AI system and by explicitly explaining these decisions to people (Miller, 2019).

DEFINITION OF XAI

In a previous publication, we integrated the different goals of XAI into the following definition: 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

(Van den Berg & Kuiper, 2020)

There is a lot of interest in XAI in academia. A systematic literature review indicates that the number of XAI publications has exploded in two years from 186 papers in 2018 to 1505 papers in 2020 (Islam et al., 2022). However, the practical application of XAI is lagging. There are relatively few papers that address the practical application of XAI. For this literature study, a conscious search was made for papers that deal with the application of XAI. However, most application papers are theoretical in nature and not based on practice. Only Dhanorkar et al. (2021) specifically discuss how XAI is implemented in practice.

How XAI relates to AI is illustrated in Figure 1 (Van den Berg & Kuiper, 2020). Figure 1 shows that an XAI system can be an integrated part of an AI system, but also a separate solution next to the AI system. In the latter case, a so-called post-hoc XAI technique is applied. Two of these techniques are SHAP (Lundberg & Lee, 2017) and LIME (Ribeiro et al., 2016). These techniques are more often applied, partly because AI models are becoming increasingly complex and are not inherently interpretable, with the result that it is not possible to explain the results of an AI system and people must rely on separate XAI systems in which post-hoc XAI techniques are applied. In short, there are two ways to generate explanations: directly from the AI system itself or using a separate XAI system.

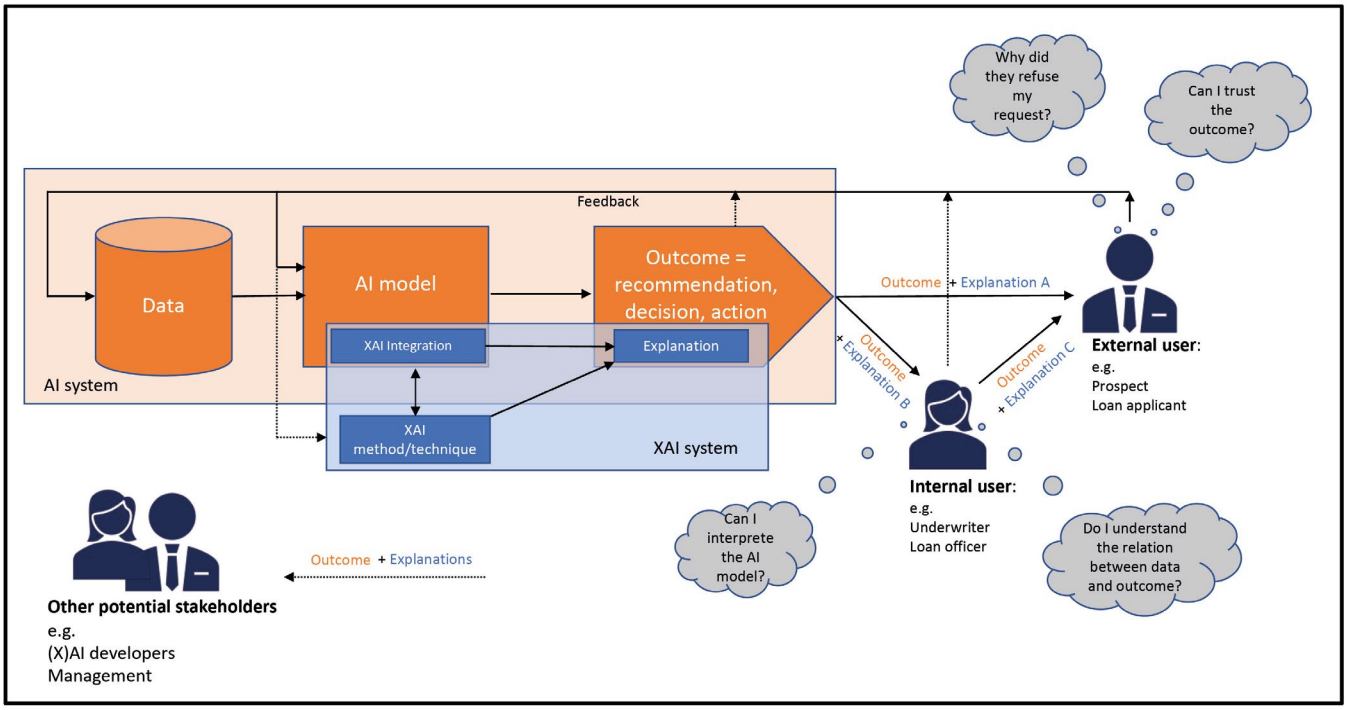


Figure 1. XAI system in relationship to AI system (based on Van den Berg & Kuiper, 2020).

In Figure 1 we use the following terms:

- AI system: A system that uses an AI model to make a decision or prediction.
- AI model: A model that is developed and used in an AI system such as a decision tree, random forest, or neural network.
- XAI system: A system to generate explanations from data and communicate these explanations to relevant stakeholders. The XAI system can be developed as part of the AI system (XAI integration) or developed with a dedicated XAI method/technique next to the AI system.
- Explainability: Property of an XAI system, i.c., the ability to explain both the technical processes of an AI system and the related human decisions.

Figure 1 also shows different stakeholders. Internal users are employees such as underwriters and loan officers. These internal users often act as a switching point (human-in-the-loop) in communication with external users such as prospects and loan applicants. The importance of these internal users in the context of AI applications is increasing. After all, the EU AI Act requires that "human oversight" is mandatory for certain riskier AI systems (EC, 2021). This means that natural persons can oversee the operation of an AI system. It also includes that these people (the internal users) must understand the outcome of the AI system.

XAI publications also show that explanations are not only necessary to provide insight into how an AI system works, called model transparency, but also into the process by which that AI system is built and implemented, called process transparency (ICO, 2019). Regulators in particular, need insight into both (Kuiper et al., 2021).

2.2 Application of XAI in the financial sector

The responsible use of AI is considered of great importance for the financial sector (FSB, 2017; Van der Burgt 2019; McWaters, 2019; EIOPA, 2021). One of the key components for the responsible use of AI is the explainability of AI systems (Bracke, 2019; Dupont, 2020). XAI is positioned as a tool to maintain and/or gain trust, which is crucial for a sector in which citizens and businesses either deposit and/or borrow money (McWaters, 2019; Van der Burgt, 2019; EBA, 2020). It is stated that the need for XAI and, accordingly, the extent to which it is used, depends on the impact and thus the proportionality of the use case: the higher the impact of the use case, the greater the need for explainability. EIOPA (2021) defines impact as the probability of harm that can be caused to an individual or organisation. Van der Burgt (2019) relates impact to the combination of materiality (both for business continuity and for consumers) and the objective of using AI in the decision-making process. This objective is mainly about how AI is deployed and can range from being descriptive (AI to analyze something that has already happened) to automated (AI that decides without human intervention). The EU AI Act applies regulations that also depend on the risk classification of a use case (EC, 2021). Significantly more and stricter rules apply to a high-risk use case than to a low-risk use case. The same is true of the level to which AI systems should be explainable in such a use case. The adoption of XAI depends not only on legislation but also on the extent to which an organisation sees XAI as a tool to ensure that its AI systems are used responsibly.

In an earlier white paper, we proposed a framework that relates types of stakeholders in the financial sector to types of explanations (Van den Berg & Kuiper, 2020). The framework illustrates that in the financial sector, relatively many different stakeholders expect different types of explanations from AI systems, ranging from explanations of the model itself, to its

outcomes, to the processes involved. Regulators in particular need insight into different types of explanations to ensure that financial service providers have control over their business operations (Kuiper et al., 2021).

2.3 AI Lifecycle Models

AI lifecycle models describe processes to design, develop, and operate AI systems and particularly machine learning (ML) systems. An AI lifecycle model provides a methodology and good practice for the execution of AI and ML projects (Martinez-Plumed et al., 2019). As part of this research project, we want to link XAI aspects to phases of the AI lifecycle. By doing this, XAI can be operationalized, i.e., XAI becomes an integral part of the process in which an AI system is designed, developed, and operated. The question for this research is then which AI lifecycle model to choose.

CRISP-DM is the de facto standard AI lifecycle model (Martinez-Plumed et al., 2019; Studer et al., 2021, Haakman et al., 2021). CRISP-DM has its origins in the nineties and was created as a standard process model for data mining projects. Recent studies identified different shortcomings. Studer et al. (2021) argue that "CRISP-DM focuses on data mining and does not cover the application scenario of ML models inferring real-time decisions over a long period". Another shortcoming identified by Studer et al. (2021) is that "CRISP-DM lacks guidance on quality assurance methodology". Haakman et al. (2021) demonstrated that traditional machine learning lifecycle models such as CRISP-DM "are missing essential steps, such as feasibility study, documentation, model evaluation, and model monitoring".

Based on these shortcomings, we choose CRISP-ML(Q) as the AI lifecycle model to link XAI aspects. CRISP-ML(Q)[2] is based on CRISP-DM, addresses the shortcomings, and consists of the following phases (Studer et al., 2021):

- Business & Data Understanding: The initial phase is concerned with tasks to define the business objectives and translate them to ML objectives, collect and verify the data quality, and finally assess the project feasibility.
- Data Engineering or Data Preparation: Building on the experience from the preceding data understanding phase, data preparation serves the purpose of producing a dataset for the subsequent modelling phase. However, data preparation is not a static phase and backtracking circles from later phases are necessary if, for example, the modelling phase or the deployment phase reveals erroneous data.
- ML Model Engineering or Modeling. The choice of modelling techniques depends on the ML and the business objectives, the data, and the boundary conditions of the project to the ML application contributing. The requirements and constraints that have been defined in business & data understanding are used as inputs to guide the model selection to a subset of appropriate models. The goal of the modelling phase is to craft one or multiple models that satisfy the given constraints and requirements.
- ML Model Evaluation. During this phase, the performance of the trained model needs to be validated on a test set.
- ML Model Deployment. The deployment phase of an ML model is characterized by its practical use in the designated field of application.
- ML Model Monitoring and Maintenance. With the expansion from knowledge discovery to data-driven applications to inferring real-time decisions, ML models are used over a long period and have a lifecycle which must be managed. The risk of not maintaining the model is the degradation of the performance over time which leads to false predictions and could cause errors in subsequent systems.

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

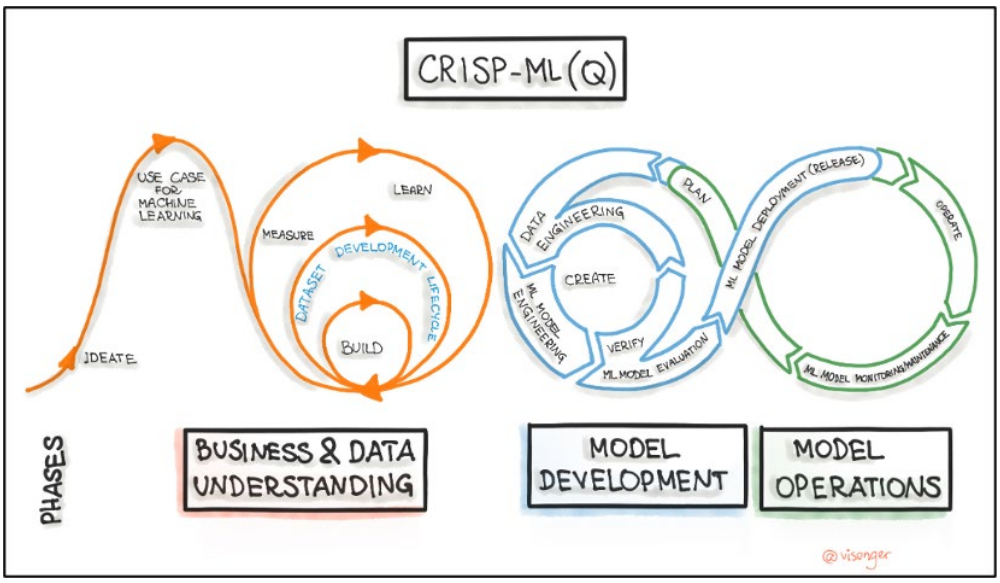


Figure 2. Machine Learning Development Lifecycle Process (https://ml-ops.org/content/crisp-ml)

3. RESEARCH APPROACH

The goal of this research is to identify aspects that play a role in implementing the explainability of AI systems in the Dutch financial sector. Aspects are considerations or choices that need to be made throughout the AI lifecycle regarding XAI to ensure that stakeholders ultimately receive a proper and meaningful explanation in case an AI system is being used. To identify the aspects, a literature study and a field study were conducted.

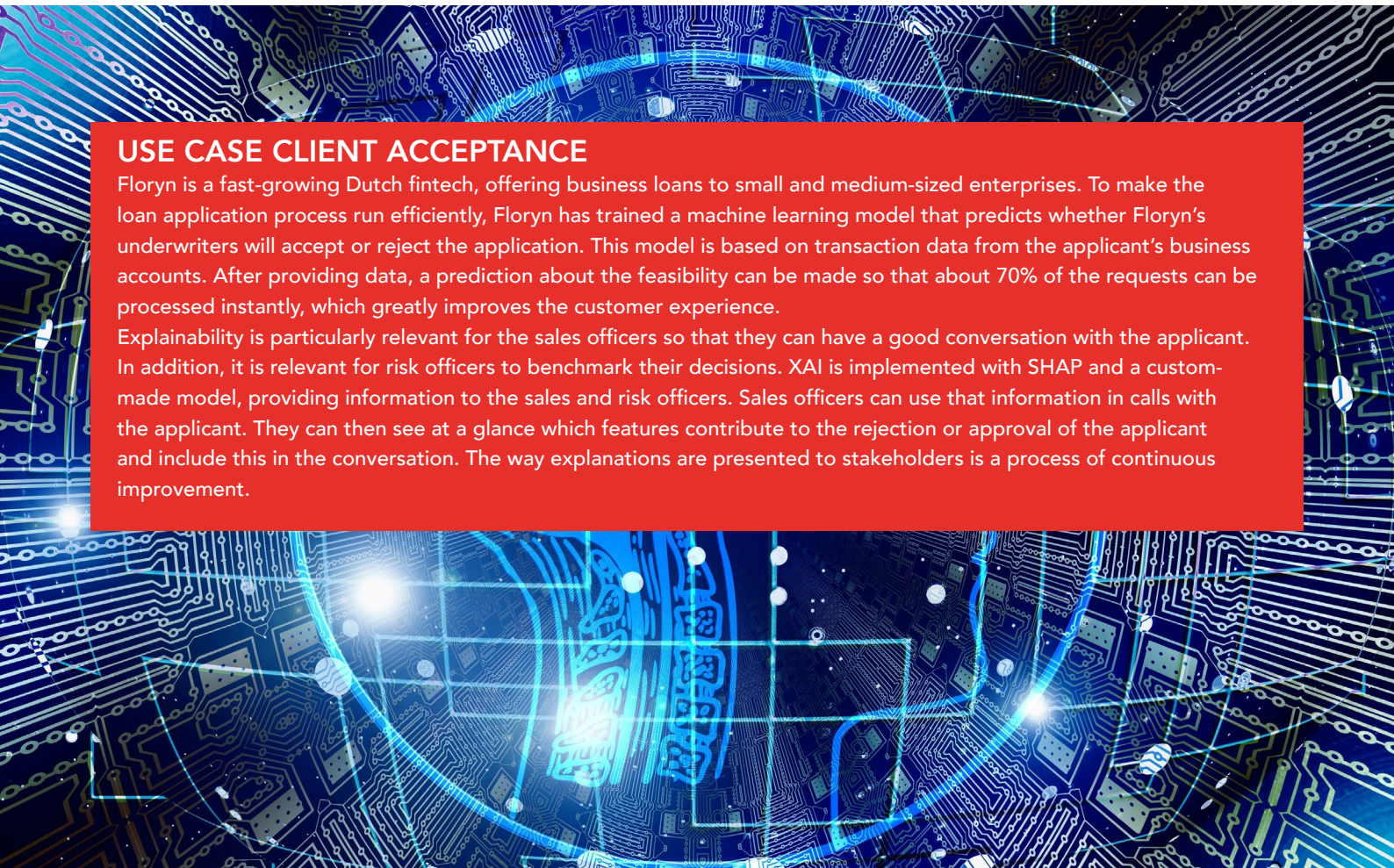
The literature study was conducted from January 2022 to April 2022. We searched for literature on the AI lifecycle and literature on XAI implementation, in general, and in the financial sector.

First, we searched for literature on the AI lifecycle. The goal of this search was to find a suitable AI lifecycle model to connect XAI-related aspects. As a result of this step, we chose CRISP ML(Q) as the AI lifecycle model. CRISP ML(Q) is explained in Section 2 of this white paper.

Second, a search was made for literature on how to implement XAI in general. Since XAI is one of the principles of ethical AI we also searched for how to implement ethical AI. The following search terms were used: 'implementation of explainable artificial intelligence (XAI)', 'how to implement explainable artificial intelligence (XAI)', 'XAI in practice', 'design patterns XAI', 'implementation of ethical artificial intelligence', and 'how to implement ethical artificial intelligence'. In addition to scientific literature, relevant literature from McKinsey and Gartner was included. The search also included snowballing (search via references of found articles). As a result, we had 32 papers that were used to extract aspects. After studying the papers, we identified 75 different aspects which we have grouped into eight categories.

Third, we searched for literature on how XAI can be implemented in the financial sector. We began this search with papers which had been previously studied by us. These papers were part of the project in which we created a framework that relates types of stakeholders in the financial sector to types of explanations (Van den Berg & Kuiper, 2020). Through snowballing, we eventually found 19 papers that we used to extract aspects. From these papers, 35 different aspects were identified and clustered into eight categories.

In the field study, we investigated four use cases. Two from Floryn: client acceptance and client review. And two from De Volksbank: arrears management and personal finance (in the mobile banking environment). One of the use cases is described in more detail.



USE CASE CLIENT ACCEPTANCE

Floryn is a fast-growing Dutch fintech, offering business loans to small and medium-sized enterprises. To make the loan application process run efficiently, Floryn has trained a machine learning model that predicts whether Floryn's underwriters will accept or reject the application. This model is based on transaction data from the applicant's business accounts. After providing data, a prediction about the feasibility can be made so that about 70% of the requests can be processed instantly, which greatly improves the customer experience.

Explainability is particularly relevant for the sales officers so that they can have a good conversation with the applicant. In addition, it is relevant for risk officers to benchmark their decisions. XAI is implemented with SHAP and a custom-made model, providing information to the sales and risk officers. Sales officers can use that information in calls with the applicant. They can then see at a glance which features contribute to the rejection or approval of the applicant and include this in the conversation. The way explanations are presented to stakeholders is a process of continuous improvement.

The use cases were studied through semi-structured interviews with employees involved in the design, development, and operation of AI in these use cases. A total of 15 interviews were held with data scientists, team leaders, model owners, product owners, risk officers, risk & compliance managers, and senior managers from both Floryn and De Volksbank. These interviews discussed how the AI system was developed and what the choices and considerations were regarding XAI and explainability. The interviews were transcribed and coded in ATLAS.ti. The codes we used while coding were the aspects we extracted from the literature. After coding and analysing the interviews, we had 254 quotes that were clustered into 41 aspects. For each aspect, we counted the total number of times the aspect was mentioned in the interviews. At the end of the field study, we had a list of aspects sorted by the number of times an aspect was mentioned in all interviews.

A workshop was the next step in this study. Here, the representatives of Floryn, Researchable and De Volksbank discussed the aspects extracted from the literature and field study. The workshop confirmed these aspects.
Based on these aspects, we developed a conceptual model and a checklist. The conceptual model contains the main aspects. These main aspects also function as a checkpoint in the checklist. The checklist was elaborated with related questions per checkpoint based on the more granular aspects. We also added organisational roles to the checkpoints indicating who is responsible and accountable to consider the checkpoint. Finally, we linked the different checkpoints to the stages of the CRISP-ML(Q) AI lifecycle model.
As a final research step, the conceptual model and checklist were discussed with two confirmatory focus groups with a total of 19 participants. Regarding the checklist, we asked the participants why they find the checklist appealing and what improvements they suggest. The feedback from the focus groups has been incorporated into the checklist discussed in Section 4. The use of the checklist is described in Section 5 of this white paper.

4. CHECKLIST

The checklist contains checkpoints with related questions to be considered when making XAI-related choices at different stages of the AI lifecycle. The target audience of the checklist consists of all roles involved in the design, development, and operation of AI and XAI systems. The goal of the checklist is to give designers and developers of AI systems a tool to ensure the AI system is developed to give a proper and meaningful explanation to each stakeholder.

Before discussing the checklist, we present the categories we used to cluster aspects. Table 1 contains these categories. A distinction is made between the organisational level and the use case level. The organisational level refers to types of aspects that need attention at the organisational level, while the other categories refer to types of aspects that are relevant at the use case level.

Table 1. Categories of XAI aspects.

Category	Meaning	Level
Overall XAI	General policies, principles, and ways of working on XAI	Organisation
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

Table 2 presents the checklist of aspects to consider at the organisational level. The greater the number of AI systems and the impact of these systems, the higher the need for guidance on dealing with explainability and XAI at the organisational level. This guidance can become concrete in artefacts such as principles and guidelines for explainability and XAI systems. These principles and guidelines should then be applied at the use case level.

Table 2. Checklist with aspects to consider on the organisational level.

Category	Checkpoint	Questions
Overall XAI	Check principles for how to deal with XAI and explainability	What are the values for how to deal with explainability and XAI?
		What are the principles for how to deal with explainability and XAI?
	Check how to design, deliver, and evaluate XAI systems	How to elicit and document explainability requirements and XAI design decisions?
		How to design, develop, operate, and evaluate XAI systems (e.g., guidelines)?
		How to define and manage the risks of XAI systems?
		How to evaluate AI models in terms of explainability and transparency?
		How to design and deliver explanations (including aspects such as human reasoning, and human-machine involvement)?
		What are the applicable laws and regulations that need to be considered in the design of XAI systems?

Table 3 shows the checklist of aspects to be considered at the use case level. The checklist is a tool for making choices and decisions, and not overlooking aspects for which those choices and decisions should be considered. In Section 5 we discuss how to use the checklist.

Table 3. Checklist with aspects to consider on the use case level.

Category	Checkpoint	Questions
Explainability and transparency in the use case	Check the level of transparency and explainability	What is the required level of transparency of the AI model?
		What is the required level of explainability of the AI model?
		What is the trade-off between the explainability and performance of the AI model?
		What is the trade-off between explainability and other requirements of the AI model such as security and intellectual property?
AI in the use case	Check the goal of the AI system	What is the purpose of AI in the use case?
		What laws and regulations apply to the use case?
		What principles and guidelines apply to the use case?
	Check the stakeholders of the AI system	Which stakeholder groups are interfacing with the product/service for which the AI system is used?
	Check the risks of the AI system	What are the risks of the AI system?
		What is the potential harm of the AI system?
		What are the ethical concerns regarding the AI system?
	Check the type of the AI model	What is the impact of the explainability requirements on the type of AI model?
		What are the benefits and costs of different types of AI models?
		What is the preferred type of AI model and why?
	Check the data of the AI system	What is the impact of the explainability requirements on the data used to train and test the AI system?
		What is the quality of the data that is used to train and test the AI system?
		What are the variables to include in the AI model?
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?
XAI system in use case	Check the goal of the XAI system	What is the purpose of XAI in the use case?
		What are the reasons to explain the AI model?
		What is the explanatory strategy (e.g., internal explanation, external or post-hoc explanation, counterfactual explanation)?
		What are the required capabilities of XAI methods and techniques?
	Check the risks of the XAI system	What are the risks of the XAI system?
		What is the potential harm of (not) providing explanations?

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?
XAI methods and techniques in the use case	Check XAI method	What logical method(s) to use to generate explanations (e.g., post-hoc explain local feature importance, ante-hoc explain global working)?
	Check XAI technique	What technical method(s) to use to generate explanations (e.g., Shap, Lime, Anchors)?
	Check XAI tool	What tool(s) to use to generate explanations (e.g., Python library, IBM AI explainability 360)?
Methods and techniques to evaluate XAI in the use case	Check how to evaluate the XAI system	What are the evaluation measures of the XAI system (e.g., user mental model, usefulness and satisfaction, model performance)?
		What method(s) to use to evaluate the XAI system (e.g., application grounded, human grounded, functionally grounded)?
		How to measure stakeholder satisfaction with the explanations provided (e.g., user engagement, Likert scale questionnaires, simulated experiments)?

The audience for this checklist consists of organisational roles involved in the design, development, and operation of AI systems, such as business analysts, model developers, machine learning engineers, data scientists, model owners, product owners, model validators, AI management, and senior management. In Table 4, we have assigned organisational roles to each checkpoint in the checklist: one responsible role and one accountable role. The responsible role is the role that performs the task and makes the choice/decision, the accountable role is the role that is ultimately responsible and approves the choice/decision.

We included the following roles in table 4:

- AI management: Runs the AI development.
- Senior management: Runs the company.
- AI process owner: Owns the AI lifecycle, i.e., the process of how AI and XAI will be developed.
- Business analyst: Analyses the requirements of the AI system and XAI system.
- Model developer: Develops the AI system and XAI system.
- Model owner: Owns the AI model.
- Product owner: Owns the AI system and XAI system.
- Model validator: Validates the AI model.

It is important to note that roles vary from organisation to organisation. For example, two roles may be combined into one, such as the role of model and product owner or the role of business analyst and model developer. Also, note that in the field of AI development there are several overlapping job titles, such as model developer, machine learning engineer, and data scientist. Finally, AI development is more and more frequently taking place in agile teams, where responsibility lies with a team rather than a single role.

Table 4 also includes a reference to the stage in the AI lifecycle in which a particular checkpoint is relevant. The AI lifecycle is based on CRISP-ML(Q) where BDU stands for business & data understanding, MD for model development, and MO for model operations. AI lifecycles can vary from one organisation to another.

Table 4. Checkpoints plotted on AI lifecycle and accountability.

Category	Checkpoint	AI Lifecycle	Responsible	Accountable
Overall XAI	Check principles for how to deal with XAI and explainability	Overall	AI management	Senior management
	Check how to design, deliver, and evaluate XAI systems	Overall	AI process owner	AI management
Explainability and transparency in the use case	Check the level of transparency and explainability	BDU	Business analyst, Model developer	Model owner
AI in the use case	Check the goal of the AI system	BDU	Business analyst	Product owner
	Check the stakeholders of the AI system	BDU	Business analyst	Product owner
	Check the risks of the AI system	BDU	Business analyst	Product owner
	Check the type of the AI model	MD	Model developer	Model owner
	Check the data of the AI system	MD	Model developer	Model owner
Stakeholder's need for explanations in the use case	Check the stakeholders of the XAI system	MD	Model developer	Model owner
	Check the stakeholder need for explanations	MD	Model developer	Model owner
XAI system in use case	Check the goal of the XAI system	MD	Model developer	Model owner
	Check the risks of the XAI system	MD	Business analyst	Product owner
Explanations in use case	Check what to explain to whom	MD	Business analyst	Product owner
	Check how to deliver the explanation	MD	Model developer	Model owner, product owner
XAI methods and techniques in the use case	Check XAI method	MD	Model developer	Model owner
	Check XAI technique	MD	Model developer	Model owner
	Check XAI tool	MD	Model developer	Model owner
Methods and techniques to evaluate XAI in the use case	Check how to evaluate the XAI system	MO	Model validator	Product owner

5. HOW TO USE THE CHECKLIST?

The checklist is a decision support tool. It contains aspects for which choices and decisions should be considered, and not overlook aspects for which choices and decisions are necessary. In the remainder of this Section, we discuss how to make use of the checklist through questions and answers. This Section is mainly based on remarks and questions raised by the focus groups.

Question:
Is the checklist meant to be a list of questions to tick off?
Answer:
In general, there are two types of checklists: read-do and do-confirm (Gawande, 2010). In the case of a read-do checklist, "one reads the item and then goes to do what's specified". A do-confirm checklist is "where you confirm you've carried out the action specified". Our checklist is meant as a do-confirm type of checklist. It contains topics that are relevant to the design, development, and operation of XAI. It is not intended as a list of questions to tick off. Our checklist is intended as a list of reminders for professionals in the form of questions that must be answered in the development of an AI system for a use case in which explainability and XAI are relevant.

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 organisation'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 organisation 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 organisational roles can be aligned with the roles used in the organisation or team. Next to that, the checkpoints and questions can be aligned with the terminology of the organisation or team.
Furthermore, the checkpoints are helpful as a guide to document the decisions related to XAI. A 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:
When should I use the checklist?
Answer:
We suggest discussing all checkpoints and related questions at the start of an AI initiative for a particular use case. Some of the checkpoints, especially those at a later stage in the AI lifecycle, are not yet relevant, but we advise you to at least try to understand these checkpoints and related questions. This creates awareness.
Furthermore, we suggest discussing the appropriate checkpoints at the start of a new stage in the AI lifecycle. Any choices and decisions that are made along the AI lifecycle based on the checkpoints and related questions should be documented accordingly.
A checkpoint is linked to only one stage of the AI lifecycle. Depending on how XAI is or will be integrated, the same checkpoint may become relevant in more stages of the AI lifecycle. And depending on the number of iterations a checkpoint may be considered multiple times in the same stage of the AI lifecycle.

Question:
Who should use the checklist?
Answer:
The roles that are involved in considering the checkpoints and making choices and decisions regarding XAI are mentioned in Table 4. As noted, these roles may differ from organisation to organisation.

Question:
Who should be involved in making XAI-related choices and decisions?
Answer:
The organisational roles or stakeholders that are necessary in the discussion of the checkpoints and questions will differ per organisation and even per use case. In general, those stakeholders should be involved that have a concern, requirement, or answer to the questions.

Question:
What is the best way to start using the checklist?
Answer:
The best way to start is to test the checklist in one of the AI projects. Based on the findings, the checklist may be improved and/or integrated into the organisation's AI lifecycle.

Question:
What is the added value of a checklist?
Answer:
Checklists are powerful business tools in which knowledge is concentrated (Gawande, 2010). According to Gawande (2010), "checklists are especially appropriate in case of increasing complexity. They make sure that knowledge is applied correctly. It prevents failures of ineptitude: this is a situation where knowledge exists, but we fail to apply it correctly. Eptitude is making sure we apply the knowledge we have consistently and correctly". It is recommended to read the Checklist Manifesto (Gawande, 2010). This book provides insights into the usefulness and added value of checklists.

Question:
Can I apply XAI without a checklist?
Answer:
Of course, you can, but remember that the application of AI is complex. And that also applies to the application of XAI. In our research, we concluded that XAI is not an afterthought. It requires many different aspects to consider, and that is the reason we developed this checklist.

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.

Question:
Does the checklist apply to both XAI integration and XAI as a separate method/technique (Figure 1)?
Answer:
Yes, it does. However, the checkpoints may apply to different stages of the AI lifecycle. When XAI is part of the AI system (XAI integration), it is conceivable that checkpoints may be discussed earlier compared to the situation that XAI is a separate method/technique (post hoc XAI). However, we argue that XAI is not an afterthought, so we recommend considering the checkpoints as early as possible.

Question:
Does the checklist apply to use cases where an AI system is procured?
Answer:
Yes, it does. The importance of explainability and XAI for a particular use case does not depend on whether an AI system is procured externally or developed internally. We recommend including explainability and XAI requirements in the procurement process. The checklist is useful to elicit these requirements.

Question:
This is not the first AI-related checklist that is proposed. What is the relation with other AI-related checklists and assessments?
Answer:
The Dutch government released an impact assessment for human rights and algorithms[3] (IAMA). One of the topics of IAMA is transparency and explainability. Explainability is discussed in general terms. References are made to the ethical guidelines from the EU (HLEG, 2019) and the "Toetsingskader Algoritmes" from the Algemene Rekenkamer[4]. IAMA, the ethical guidelines from the EU, and the "Toetsingskader Algoritmes" have a much broader scope than our checklist and discuss explainability as one of the requirements for ethical and responsible AI. Explainability is discussed and translated into requirements on a high level. Our checklist is more comprehensive and provides more guidance on how to deal with explainability and XAI.
Koster et al. (2021) published a checklist for explainable AI in the insurance domain. This paper was included in our literature study and was used to extract aspects that were the basis for our checklist. Compared to Koster et al. (2021) our checklist is more comprehensive and links the checkpoints to stages of the AI lifecycle.

3 https://www.rijksoverheid.nl/documenten/rapporten/2021/02/25/impact-assessment-mensenrechten-en-algoritmes
4 https://www.rekenkamer.nl/onderwerpen/algoritmes-digitaal-toetsingskader

6. CONCLUSION AND CALL TO ACTION

Both theory and practice show that XAI is more than just making sure the AI application is explicable. There are many aspects to consider when applying XAI in a use case. The extent to which these aspects need to be considered depends on the impact of the use case. As a rule of thumb, the greater the impact of the use case, the more important requirements such as explainability and transparency, and the more relevant to make thoroughly informed choices and decisions about XAI.

To guide all organisational roles involved in the design, development, and operation of AI and XAI systems, we have created a checklist. This checklist contains checkpoints and related questions that should be considered to make XAI-related choices at different stages of the AI lifecycle. The goal of the checklist is to enable a situation where stakeholders of an AI system receive a proper and meaningful explanation. Using the checklist also requires guidance. This guidance is provided in the form of questions and answers.

This project taught us that explainable AI is still in its infancy. There are still many areas that need further research and improvement. The checkpoints in the checklist are essentially those areas. The area that stands out is "Check how you provide the explanation". The field study showed that this is the most urgent area for future research. With a technique like SHAP it is now possible for a data scientist to find out the most important explanatory features, but communicating this information in an understandable way to internal users turns out to be a challenge. We have planned further research in this area and have applied for a new project. The research question of this project will be: "In what ways can a meaningful explanation be generated and communicated to internal users of an AI system within the financial sector and how can it be assessed whether that explanation meets the requirements of these users and applicable laws and regulations?".

We end this paper with a call to action. As with all tools, the proof of the pudding is in the eating. Please try the checklist in your practice and let us know your experiences and points for improvement. We wish you every success in applying the checklist.

REFERENCES

- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE access, 6, 52138-52160.
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion, 58, 82-115.
- Bauer, K., Hinz, O., Aalst, W., & Weinhardt, C. (2021). Expl (AI) n It to Me–Explainable AI and Information Systems Research. Business & Information Systems Engineering: Vol. 63, No. 2.
- Bracke, P., Datta, A., Jung, C., & Sen, S. (2019). Machine learning explainability in finance: an application to default risk analysis. Staff Working Paper No. 816. London, United Kingdom: Bank of England.
- Dhanorkar, S., Wolf, C. T., Qian, K., Xu, A., Popa, L., & Li, Y. (2021). Who needs to know what, when? Broadening the Explainable AI (XAI) Design Space by Looking at Explanations Across the AI Lifecycle. In Designing Interactive Systems Conference 2021 (pp. 1591-1602).
- Dupont, L., Fliche, O., & Yang, S. (2020). Governance of artificial intelligence in finance. Banque De France.
- Dutch Digital Delta. (2019). Sleuteltechnologieën. Retrieved from https://dutchdigitaldelta.nl/sleuteltechnologieen
- EBA (European Banking Authority). (2020). Report on big data and advanced analytics. Retrieved from https://www.eba.europa.eu/sites/default/documents/files/document_library/Final%20Report%20on%20Big%20Data%20and%20Advanced%20Analytics.pdf
- EC (European Commission). (2021). Proposal for a Regulation of the European Parliament and of the Council, Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending certain Union Legislative Acts. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206
- EIOPA (European Insurance and Occupational Pensions Authority). 2021. AI Governance Principles towards Ethical and Trustworthy AI in the European Insurance Sector. Retrieved from https://www.eiopa.europa.eu/sites/default/files/publications/reports/eiopa-ai-governance-principles-june-2021.pdf
- Gawande, A. (2010). Checklist manifesto, the (HB). Penguin Books India.
- Gerlings, J., Shollo, A., & Constantiou, I. (2020). Reviewing the Need for Explainable Artificial Intelligence (XAI). arXiv preprint arXiv:2012.01007.
- Haakman, M., Cruz, L., Huijgens, H., & van Deursen, A. (2021). AI lifecycle models need to be revised. Empirical Software Engineering, 26(5), 1-29.
- HLEG (The High-Level Expert Group on Artificial Intelligence) (2019). Ethics Guidelines for Trustworthy AI. EU Document. Retrieved from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
- ING. (2020). AI vindt zijn weg naar alle sectoren. Retrieved from https://www.ing.nl/media/ING_EBZ_ai-vindt-zijn-weg-naar-alle-sectoren_tcm162-209612.pdf
- Islam, M. R., Ahmed, M. U., Barua, S., & Begum, S. (2022). A systematic review of explainable artificial intelligence in terms of different application domains and tasks. Applied Sciences, 12(3), 1353.
- Kemper, J., & Kolkman, D. (2019). Transparent to whom? No algorithmic accountability without a critical audience. Information, Communication & Society, 22(14), 2081-2096.
- Koster, O., Kosman, R., & Visser, J. (2021, September). A checklist for explainable AI in the insurance domain. In International Conference on the Quality of Information and Communications Technology (pp. 446-456). Springer, Cham.
- Kuiper, O., Berg, M. van den., Burgt, J. van der, & Leijnen, S. (2021). Exploring explainable AI in the financial sector: perspectives of banks and supervisory authorities. In Benelux Conference on Artificial Intelligence (pp. 105-119). Springer, Cham.
- Liao, Q. V., & Varshney, K. R. (2021). Human-centered explainable ai (xai): From algorithms to user experiences. arXiv preprint arXiv:2110.10790.
- Lundberg, S., & Lee, S. I. (2017). A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874.
- Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J. H., Kull, M., Lachiche, N., Ramirez-Quintana, M.J. & Flach, P. A. (2019). CRISP-DM twenty years later: From data mining processes to data science trajectories. IEEE Transactions on Knowledge and Data Engineering.
- McWaters, R. J. (2019). Navigating Uncharted Waters: A Roadmap to Responsible Innovation with AI in Financial Services: Part of the Future of Financial Services Series. World Economic Forum.
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial intelligence, 267, 1-38.
- Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2021). From

what to how: an initial review of publicly available AI ethics tools, methods, and research to translate principles into practices. In Ethics, Governance, and Policies in Artificial Intelligence (pp. 153-183). Springer, Cham.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).
- Scientific Council for Government Policy (Wetenschappelijke Raad voor het Regeringsbeleid) (2021) Opgave AI. De nieuwe systeemtechnologie, WRR-Rapport 105, Den Haag: WRR.
- Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., & Müller, K. R. (2021). Towards CRISP-ML (Q): a machine learning process model with quality assurance methodology. Machine Learning and Knowledge Extraction, 3(2), 392-413.
- Van den Berg, M., & Kuiper, O. (2020). XAI in the Financial Sector. Retrieved from https://www.internationalhu.com/research/projects/explainable-ai-in-the-financial-sector
- Van der Burgt, J. (2019). General Principles for the use of AI in the Financial Sector. Retrieved from https://www.dnb.nl/actueel/algemeen-nieuws/dnbulletin-2019/dnb-komt-met-richtlijnen-voor-gebruik-kunstmatige-intelligentie/