# Identifying XAI User Needs: Gaps between Literature and Use Cases in the Financial Sector

Jenia Kim[1,*,†], Henry Maathuis[1,2,*,†], Kees van Montfort[3] and Danielle Sent[1,2]

[1]HU University of Applied Sciences Utrecht, Research Group Artificial Intelligence, Heidelberglaan 15, 3584 CS Utrecht, The Netherlands
[2]Jheronimus Academy of Data Science, Tilburg University, Eindhoven University of Technology, St. Janssingel 92, 5211 DA 's-Hertogenbosch, The Netherlands
[3]Amsterdam University of Applied Sciences, Wibautstraat 2-4 1091 GM Amsterdam, The Netherlands

## Abstract

One aspect of a responsible application of Artificial Intelligence (AI) is ensuring that the operation and outputs of an AI system are understandable for non-technical users, who need to consider its recommendations in their decision making. The importance of explainable AI (XAI) is widely acknowledged; however, its practical implementation is not straightforward. In particular, it is still unclear what the requirements are of non-technical users from explanations, i.e. what makes an explanation *meaningful*. In this paper, we synthesize insights on meaningful explanations from a literature study and two use cases in the financial sector. We identified 30 components of meaningfulness in XAI literature. In addition, we report three themes associated with explanation needs that were central to the users in our use cases, but are not prominently described in literature: *actionability*, *coherent narratives* and *context*. Our results highlight the importance of narrowing the gap between theoretical and applied responsible AI.

## Keywords

Explainable AI, Finance, Human-Centered Evaluation

## 1. Introduction

Artificial Intelligence (AI) is increasingly being integrated into business processes of financial services providers in the Netherlands. This goes hand in hand with awareness within these organizations that AI needs to be implemented responsibly (e.g., [1]). One of the aspects of responsible application of AI is ensuring that the outcomes and the internal workings of an AI-based system are understandable for the non-technical employees who interact with it, such as risk underwriters and claim handlers (e.g., [2]). This is important since these employees need to be able to communicate the reasoning behind decisions to the customers, for example, explain why an insurance claim, or a loan request, was rejected.

While financial services companies acknowledge the importance of explainable AI (XAI), they indicate that its practical implementation is not straightforward. In particular, it is unclear

*HHAI-WS 2024: Workshops at the Third International Conference on Hybrid Human-Artificial Intelligence (HHAI), June 10—14, 2024, Malmö, Sweden*

*Corresponding author.
†These authors contributed equally.
Qi jenia.kim@hu.nl (J. Kim); henry.maathuis@hu.nl (H. Maathuis)
8 0009-0008-5067-4640 (J. Kim); 0009-0002-5542-0478 (H. Maathuis); 0009-0007-2803-5095 (K. v. Montfort); 0000-0002-4703-5345 (D. Sent)

what constitutes a “good" explanation from the point of view of the non-technical user. While the objective quality of an explanation (e.g., correctness, completeness) can be measured by the developers of the AI system, it does not ensure that the explanation is *meaningful* to the end-user and achieves goals like understandability, trust and good decision making.

In the FIN-X project[1], researchers and organizations from the financial sector work together to address this gap. We aim to develop practical guidelines that will detail (a) what is a meaningful explanation for a user of an AI system, (b) how to communicate explanations in a meaningful way, and (c) how to evaluate the meaningfulness of explanations. To achieve this goal, we synthesize insights from literature, as well as legal requirements from the GDPR and EU AI Act, and requirements from the use cases provided by our industry partners. Using this information, we create prototypes and evaluate them with the intended users.

In this paper, we report the findings from the first phases of the project. We focus on findings from literature and the use cases; the legal requirements are out of scope for this paper. We present some gaps and insights that emerged from comparing the explanation requirements mentioned in the use cases and those identified in literature. This highlights the complementary nature of academic research and practical real-world implementations, and the importance of narrowing the gap between theoretical and applied Responsible AI.

## 2. Literature Study

As part of the FIN-X project, a systematic literature review was performed, which focused on how explanations are evaluated in empirical studies with users. The idea is that the properties that researchers choose to evaluate with users are components of what is considered a meaningful explanation by the XAI research community. The systematic review on aspects of meaningful explanations is currently under review [3]; in this study, we focus on a subset of our findings.

### 2.1. Method

In November 2023, we performed a search in five databases (ACM Digital Library, Scopus, Web of Science, IEEE Xplore and PubMed) to find abstracts that contain two elements: (1) mention of *explainable AI* or *XAI*, and (2) mention of words related to evaluation from a user perspective: *meaningful*, *trustworthy*, *understandable* or *interpretable*. The query returned 3,103 papers; after deduplication, 1,655 unique papers remained. These papers went through a few rounds of filtering, after which 73 papers remained that fulfilled our inclusion criteria: papers that (a) involve an AI-based system with explanations, and (b) report an empirical evaluation of the explanations in a user study.

### 2.2. Selected insights from the literature study

We systematically collected the properties evaluated in the user studies in our set of papers; we consider these properties to be components of a meaningful explanation. We found that a meaningful explanation has 30 properties, according to the reviewed literature. We categorized these 30 properties along three dimensions, (as also shown in Figure 1):

[1] https://www.internationalhu.com/research/projects/fin-x

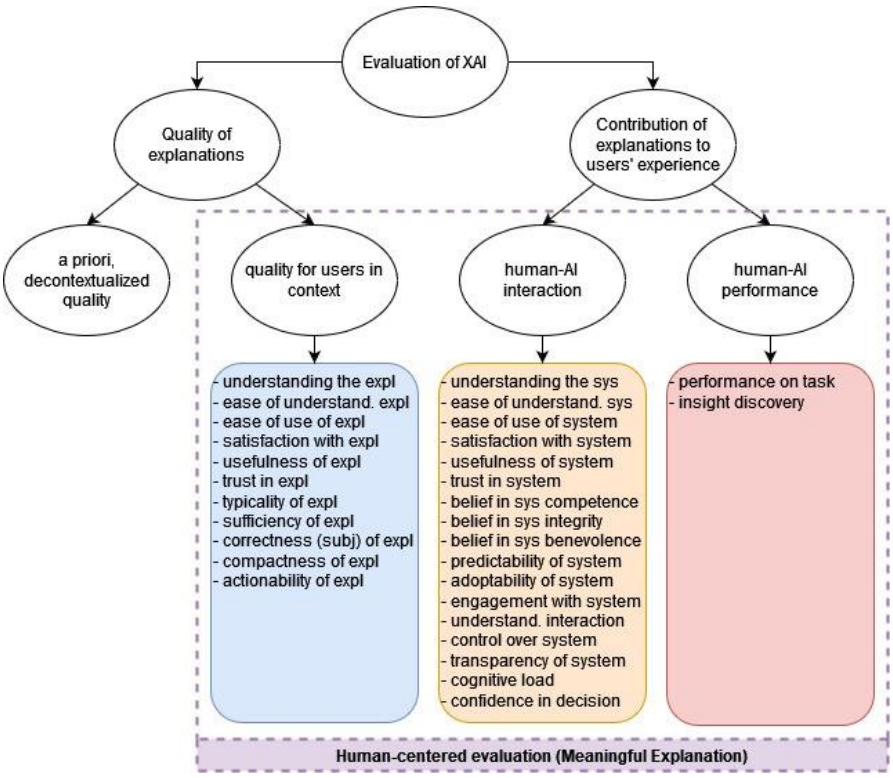


**Figure 1:** Components of a meaningful explanation (expl=explanation, sys=system)

- The **in-context quality of the explanation** (11 components). Is the explanation *satisfying*, *understandable*, *useful*, *actionable*, *sufficient*, *compact*, *trustworthy*, *correct*, *typical*, *easy to understand*, and *easy to use*?
- The contribution of the explanation to **human-AI interaction** (17 components). Does the explanation help the user to better *understand the AI system*? Does it improve the user's perception of the AI system as *trustworthy*, *useful*, *satisfying*, *competent*, *honest*, *benevolent*, *controllable*, *predictable*, *transparent*, *easy to understand*, *easy to use*, and *engaging*? Does the explanation help the user to better *understand the interaction with the AI system*? Does it make the interaction less *cognitively demanding*? Does it increase the user's *confidence in the decision*? Does it increase the *readiness to adopt the AI system* and use it?
- The contribution of the explanation to **human-AI performance** (2 components). Does the explanation improve the user's *performance on the task*? Does it help the user to *discover new insights*?

## 3. Use Cases

As part of the FIN-X project, each collaborating company was asked to submit a use case, involving an AI system that is currently used in the organization. Here, we focus on two use cases. **Company A** offers credit (loans) to businesses. Their AI application estimates the chance

of approving a credit request; it outputs a score (approval chance) and a local feature importance explanation, i.e. the three top factors contributing to the score. The AI application of **Company B** detects risk of fraud in insurance claims; it outputs a risk score and a local feature importance explanation, i.e. the five top factors contributing to the score.

We chose to focus on these two use cases because they are similar in the type of output (score) and the type of explanation (local feature importance); moreover, in both cases, non-technical employees (risk underwriters in A, claim handlers in B) make a final decision with the support of the AI system.

## 3.1. Method

In each company, five stakeholders of the AI systems were interviewed, who were indicated by the company as knowledgeable about the use case. In company A, the interviewees included non-technical users, developers and management. In company B, the interviewees included developers and consultants. The interviews were conducted in a semi-structured manner: several predefined questions were asked, but the interviewees were also encouraged to talk freely. The interviews were recorded and transcribed. The transcriptions were labeled according to themes of interest.

Since we did not want to influence interviewees' responses, we did not present them with the results of our literature study, nor ask them about the aspects of meaningfulness that we identified in literature. Aspects mentioned in the interviews were only afterwards compared to those found in literature. We focused on elements that were mentioned in the interviews, but were not prominent in literature, i.e. needs of users that are overlooked by the research community.[2]

## 3.2. Selected insights from the interviews

We focus on three insights that we qualitatively estimated as important and recurring in the interviews. Below, we shortly describe each insight and provide two illustrative quotes; the emphasis in the quotes is ours. For each quote, we mention the role of the interviewee in the company; sometimes the speakers are not users, but they talk from the perspective of non-technical users.

### Actionable explanations

Interviewees from both use cases emphasized the importance of the *actionability* of explanations. It is not enough for the users to understand the explanation and the AI's recommendation; an explanation is perceived as meaningful if it directs the user towards an action or a next step.

> *"For me, the explanation of the system is good if it is presented in a clear manner. So, digestible information that I* ***can do something with****, so to speak."* (Company A; sales person, former risk underwriter)

[2]The opposite comparison (i.e. which aspects from literature are not mentioned in the interviews) is not possible in this case, since the fact that a person does not mention something in an open-ended conversation does not mean that they do not find it important.

*"It's clear why the model made [a] decision. But the claim investigators are like, okay, but what do you want me* ***to do with this****? So the actionability part is also highly requested."* (Company B; ML engineer)

### Coherent narratives and scenarios

In the fraud detection use case (company B), interviewees expressed the need for a coherent narrative that ties the separate indicators together. It is not enough for the users to see the various factors that contribute to the AI's recommendation; an explanation is perceived as meaningful if, similarly to humans, it constructs a fraud scenario (a story) from the combination of the factors.[3]

*"What I know from the customers and everything [is] that they're more* ***missing the full scenario*** *of why this hit. Here you're just presenting different things, but it's like also trying to figure out how these correlate and what is the full picture of this."* (Company B; consultant)

*"I can imagine customers saying, like, we got this sort of scenario, but now it's reduced to* ***a set of factors in which the scenario is a bit lost****.* (Company B; ML engineer)

### Additional context

Interviewees from both use cases indicated that the output of the AI system and the explanations are a good starting point for the analysis, but they are always supplemented by additional (qualitative) information. AI recommendations and explanations are meaningful only in combination with additional context that is provided by the human expert.

*"It is a very good starting point for our analysis, the data and the bank statements, but above all, I think, powerfully* ***supplemented with some qualitative data****."* (Company A; sales person, former risk underwriter)

*"But we are looking into the* ***overall context****. [...] This additional insight, this additional information is, I think, super helpful."* (Company B; consultant)

## 4. Discussion and Conclusion

We observed that some aspects of a meaningful explanation that are important to users of real-world AI systems are not prominently featured in current XAI literature. First, the **actionability of explanations** was one of the most salient points in the interviews, but in the literature study it was found only in one paper out of the 73 included in the study.

Second, the need for **coherent narratives** to be constructed out of the individual factors contributing to the AI recommendation was central in the fraud detection use case, but it was

[3]The customers mentioned in the quotes are the companies that use the AI application provided by company B, and particularly the claim handlers, who are the users of the system.

not mentioned in the literature that we reviewed. Notably, this requirement seems to be specific to the fraud detection use case; in this decision-making process, seeing the explanation as separate factors is not meaningful, because the overall narrative (a plausible fraud scenario) is lost. In the credit approval use case, on the other hand, the need for a coherent narrative was not mentioned, probably because the process of approving a credit request is aligned with checking whether individual factors are satisfied.

Third, the role of **context**, i.e. additional (external, qualitative) knowledge to supplement AI was not directly evaluated in the literature we reviewed. However, this aspect might be related to a variable that is explored in some studies: level of expertise. Some studies found that explanations are more beneficial for people with higher expertise in the task (e.g., [4]); this could be due to the ability of these users to bring external knowledge into the task, which helps them to make sense of the explanations.

To conclude, we discovered from the literature study that a *meaningful explanation* is a complex, multi-dimensional construct. Comparing the properties discussed in the literature with those mentioned in the use cases, we observed that some aspects that are important to users are not prominent in the current XAI literature. In addition, we observed that some aspects, such as the need for coherent narratives, are use-case specific. Even though we only focused on two use cases, this already provided us with valuable insights that have not yet been described in the literature.

Our findings highlight the importance of conducting studies that evaluate XAI in an application grounded setup, i.e. evaluation with a real task and the expert intended users of the application [5]. As we report here, in a real-world setup, user needs can be discovered which might be overlooked otherwise. Future work will explore how these insights can be translated into actionable evaluation methods to be implemented in user studies.

## References


[1] M. Van den Berg, J. Gerlings, J. Kim, Empirical research on ensuring ethical ai in fraud detection of insurance claims: A field study of dutch insurers, in: European Conference on Artificial Intelligence, Springer, 2023, pp. 106–114.

[2] A. Bertrand, J. R. Eagan, W. Maxwell, Questioning the ability of feature-based explanations to empower non-experts in robo-advised financial decision-making, in: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, 2023, pp. 943–958.

[3] J. Kim, H. Maathuis, D. Sent, Human-Centered Evaluation of Explainable AI Applications: a Systematic Review (under review).

[4] B. Ghai, Q. V. Liao, Y. Zhang, R. Bellamy, K. Mueller, Explainable active learning (XAL) toward AI explanations as interfaces for machine teachers, Proceedings of the ACM on Human-Computer Interaction 4 (2021) 1–28.

[5] F. Doshi-Velez, B. Kim, Considerations for evaluation and generalization in interpretable machine learning, Explainable and interpretable models in computer vision and machine learning (2018) 3–17.