Improving Artificial Intelligence Through User Feedback in eXplainable Artificial Intelligence (XAI) Systems

Design Recommendations for Sustaining Long-Term Human-XAI Interaction in Industry 4.0

Master's thesis in Computer science and engineering

Negin Hashmati and Hugo Wärnberg

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2024

Master's thesis 2024

Improving Artificial Intelligence Through User Feedback in eXplainable Artificial Intelligence (XAI) Systems

Design Recommendations for Sustaining
Long-Term Human-XAI Interaction in Industry 4.0

NEGIN HASHMATI AND HUGO WÄRNBERG

UNIVERSITY OF
GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Division of Interaction Design
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2024

Improving Artificial Intelligence Through User Feedback in eXplainable Artificial Intelligence (XAI) Systems
Design Recommendations for Sustaining Long-Term Human-XAI Interaction in Industry 4.0
NEGIN HASHMATI AND HUGO WÄRNBERG

Academic Supervisor: Mohammad Obaid, Department of Computer Science and Engineering
Industry Supervisor: Emmanuel Brorsson, ABB
Examiner: Palle Dahlstedt, Department of Computer Science and Engineering

Master's Thesis 2024
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Visualization of a control room AI-generated with Midjourney (`https://www.midjourney.com/home`).

Typeset in LaTeX
Gothenburg, Sweden 2024

Improving Artificial Intelligence Through User Feedback in eXplainable Artificial Intelligence (XAI) Systems
Design Recommendations for Sustaining Long-Term Human-XAI Interaction in Industry 4.0
NEGIN HASHMATI AND HUGO WÄRNBERG
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

This thesis explores Human-XAI interaction and the vital role of user motivation in providing feedback to XAI systems in industrial settings, focusing on the design of Incremental Explanatory Training (IET) components that enable feedback. A well-adjusted XAI interface is necessary for feedback provision and in turn, the feedback is used to improve the accuracy of the Artificial Intelligence. Drawing from the interdisciplinary field of XAI, the industry 4.0 context, and User-Centered Design (UCD) practices, it investigates the motivations required for sustained user engagement, identifies interface components that enhance feedback provision, and offers design recommendations for optimizing the user experience and sustaining motivation for feedback provision. This research addresses three key questions: 1) the types of motivations necessary for user feedback, 2) the interface components fostering motivation, and 3) design recommendations for effective feedback mechanisms adhering to motivation. As a result of this research, it was shown that both *extrinsic* and *intrinsic* motivations are necessary to encourage sustained feedback provision and that interface components with XAI improvement statistics and effective interaction modalities are vital for fostering motivation. Eight design recommendations were identified that offer generalized insights into important considerations for IET components to contribute to sustained motivation and user engagement for feedback provision.

Keywords: motivation, feedback, explainable artificial intelligence, industry 4.0, design recommendations, human-AI interaction, interaction design, user-centered design, incremental explanatory training.

Acknowledgements

For his excellent feedback, assistance, and guidance throughout the entire project, we would like to thank **Emmanuel Brorsson**, our supervisor from ABB. We are also grateful to **Dawid Ziobro** at ABB for entrusting us with this project.
We also express our gratitude to our academic supervisor **Dr. Mohammad Obaid** for his invaluable insights, advice, and guidance in the scientific and academic writing process.
Additionally, we would like to thank **Dr. Andreas Darnell** for his incredible hospitality, support, and forthcomingness during the entire project.
Their contributions have been invaluable to the successful completion of this thesis.

Negin Hashmati and Hugo Wärnberg, Gothenburg, May 2024

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis, listed in alphabetical order:

AI	Artificial Intelligence
HCI	Human-Computer Interaction
IET	Incremental Explanatory Training
IPG	Interactive Prediction Graph
ML	Machine Learning
UCD	User-Centered Design
UI	User Interface
UX	User Experience
XAI	Explainable Artificial Intelligence

Contents

List of Figures

List of Tables

1
Introduction

While trust and mutual understanding are important for relationships between humans, it is also important for relationships between humans and Artificial Intelligence (AI) systems. For humans to be able to trust and understand AI systems, the systems need to explain how and why they act the way they do [124]. In recent years AI has found its way into several industries, including many industrial production processes. There is a need to make these AI systems transparent and understandable in these contexts, as it affects production quality and efficacy [125]. Currently, the inner workings of AI systems are seen as black boxes, where the relationship between input and output is unclear. The field of Explainable AI (XAI) is dedicated to making this black box more transparent. In recent years XAI, which is traditionally a computer science-driven field, has opened up to interdisciplinary researchers as there is a need to cater the systems to all kinds of users [28]. In industrial contexts, the success of AI is not only dependent on explainable interfaces but also that the end-users provide feedback to and interact with the model to improve it [14]. This thesis primarily explores this interaction and the motivations required to sustain it, i.e. how the users can provide feedback to the AI system to improve its performance. As suggested by previous research, a well-functioning XAI interface is necessary to enable this human-in-the-loop scenario [116]. Enabling end-users to iteratively train the AI models through feedback enables more useful AI implementations. If the AI models are not properly aligned with the goals and motivations of the end-users, its usefulness will be questioned. To ensure longevity and ongoing incremental improvements to AI in industrial contexts, the end-users must have the ability to - and feel motivated to - provide feedback to the AI systems. As interaction designers, it is crucial to fully understand the space of XAI, its functional deployment in industrial contexts, and the end-user's practical needs in that given setting. In particular, the goal of this thesis is to understand what kinds of motivation end-users require to provide feedback to the AI, design an interface thereafter, and identify design recommendations for sustaining motivation in feedback provision.

An effort to research the industrial application of AI is the *EXPLAIN Project*, an EU-funded project to create a new AI lifecycle for industry 4.0 [5]. Industry 4.0 is the term commonly used for what is considered the fourth industrial revolution, which focuses on the advancement of digitalization in factories, smart objects, and the use of internet technologies [64]. This thesis is done in collaboration with ABB as part of the EXPLAIN project, and with their partners aiming to integrate AI systems in their industries. This thesis in particular focuses on the paper production process, where an AI monitoring the KRAFT process is going to be implemented.

The KRAFT process is the production of fibers and the removal of lignin when producing paper [6], and due to the length of the delignification process (which is a part of the KRAFT process), it is beneficial to utilize AI to create a prediction of what happens when variables in the process change. In this thesis, this use case is used to investigate explainability practices suitable for the domain and different methods of enabling user feedback to the AI system.
The EXPLAIN project [5] defines a new AI lifecycle that aims at being more transparent and explainable throughout the entire lifecycle, by having end-users involved in all steps of the process. An important new part of the lifecycle is Incremental Explanatory Training (IET), where end-users get to contribute to the accuracy of the AI by providing feedback on the explanations and outputs the system gives them. Through feedback from the users, the AI models are trained and improved continuously by experts in Machine Learning (ML) [5].
To ensure that AI systems are successfully implemented in industrial contexts, well-designed IET components are necessary. Currently, there is little research covering this topic. There is, however, research suggesting there is a need for it [2, 14, 48]. Primarily, the design of the IET components is the focus of this thesis. As previously mentioned, an important consideration is the motivation to give feedback, to ensure continuous engagement with the system and improvement of the AI models. The overarching objective of this thesis is therefore to design successful IET components that support motivation for feedback provision in industrial AI applications. To address this objective, three research questions have been chosen and are described in the following section.

1.1 Research Questions

The design of IET components is crucial for the EXPLAIN project's [5] goal to successfully introduce AI in industrial contexts. Furthermore, to ensure longevity and continuous usage of the design, it needs to support user motivation in providing feedback. The identification of the need for IET components from previous research [2, 14, 48], and from ABB within the EXPLAIN project specifying the importance of motivation, solidifies the following research questions:

- RQ1. What kinds of motivations are necessary to encourage users to offer feedback to the Explainable Artificial Intelligence system throughout the lifecycle of industrial Artificial Intelligence?
- RQ2. What interface components contribute to motivating the users to offer feedback to the Explainable Artificial Intelligence system?
- RQ3. What design recommendations can be identified for providing feedback to the Explainable Artificial Intelligence system?

2
Background

An underlying motivation for the EXPLAIN project [5] is the fact that only 15% of all AI applications in industry will be successful. The adoption rate is particularly low due to the opaque nature of AI solutions [46]. The application of AI in industry differs significantly compared to requirements identified in consumer-oriented applications [47]. For instance, in industrial contexts, the users are usually expert users - commonly called operators - who are highly familiar with the complex systems used in their workspace. They monitor an industrial process from an on-site control room. In this chapter, industrial applications of AI will be presented and related to some of the previous work done in the area of XAI. This chapter will also look into the specific industrial process this research relates to.

2.1 AI in Industrial Contexts

AI in industrial contexts must be highly specified to industrial problem settings, which usually are highly optimized processes. Although errors and breakdowns are relatively rare events, when they do happen they have big consequences for production output and quality. One approach for implementing AI in industrial contexts is following a nine-stage workflow process which involves iterations and several feedback loops to create and evaluate ML models [7]. Compared to the nine-stage workflow where the ML model is created and validated by an ML expert, the EXPLAIN project argues that in industrial contexts there needs to be a way for the domain expert to use and evaluate the system so that they can further optimize the model to adhere to their workflow [5, 9]. To this end, it is important that the operators can work with the AI system while still maintaining control of the actions and decisions made. Typically in the creation of ML models for industrial contexts, domain experts are only involved in the requirement elicitation and data labeling phases, while being left out in the evaluation phase of the model [8]. However, to succeed in implementing AI applications in industry, the AI systems need to be human-centered and continually open for domain expert feedback to address the challenges and requirements of industry processes. Furthermore, since ML metrics are notoriously difficult to understand for non-ML experts, this iterative feedback process needs to be done in a user-friendly way.
In light of these issues, there is an initiative to create a new AI lifecycle for Industry 4.0. The EXPLAIN project is aimed at increasing explainability and transparency at every step of the AI life cycle [5]. In particular, this project introduces four new steps in the life cycle of industrial AI, focused on the deep involvement of

domain experts (Fig. 2.1). In the first step, called Explanatory Training, domain experts will work with ML experts to directly interact with the ML model during the training process and receive explanations of the model output. They will also be able to provide feedback. In the second step, called Explanation Review, domain experts will validate ML solutions by gaining insights into the internal reasoning of the trained model to ensure that relevant concepts are learned from the provided data. In the third and fourth steps, every output of the ML models will be explained to end-users, and they can also give feedback and trigger IET. This IET is crucial to ensure that AI in industrial contexts is aligned with the workflow of the operators and the problems they face.

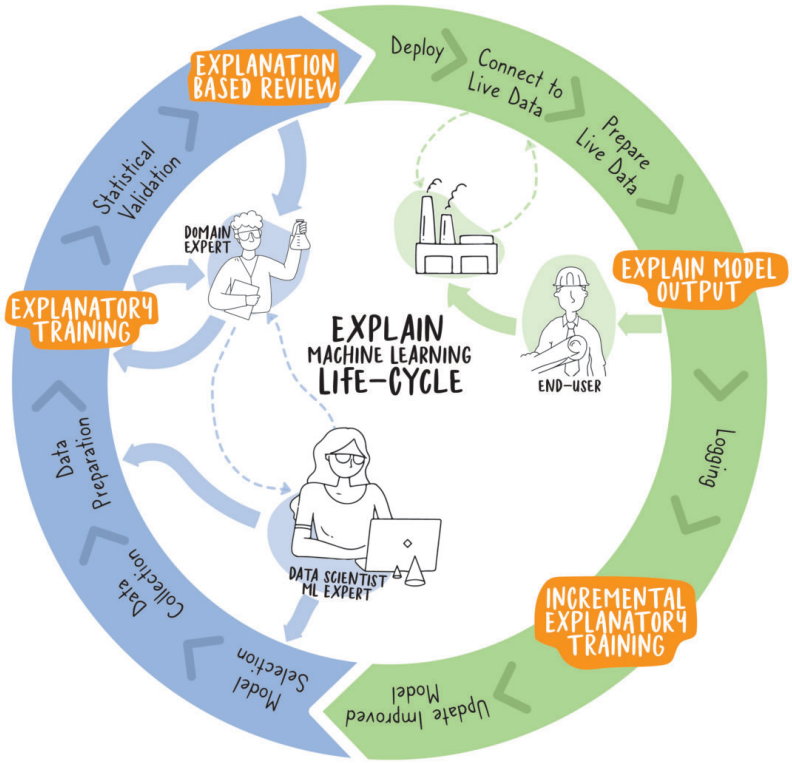


Figure 2.1: Illustration of the EXPLAIN AI life cycle from [5].

This research is conducted in collaboration with ABB, a Swedish-Swiss multinational corporation at the forefront of sustainable and productive innovations within industry. They are a part of the EXPLAIN project as one of the automation and AI providers. In this context, ABB is also involved in the interaction design for these novel AI solutions. To this end, this thesis is written to research the IET components, a crucial step in the new AI lifecycle. The aim is to identify user requirements, explore User Interface (UI) and User Experience (UX) designs that engage and motivate the users to keep improving the system, and finally evaluate the solution and suggest design recommendations for future iterations or implementations.

2.1.1 Real-life Applications of XAI

In this section, some real-life applications of XAI and how they integrate with their respective domains will be presented.

One technique of making AI explainable is demonstrated by the company FICO [86]. Their approach involves feeding different inputs into an AI system and then observing the different outputs, to make the decisions of the AI more transparent [4]. This approach has been widely discussed in the literature to increase transparency [32, 33]. For example, an insurance company using AI to generate a quote must analyze factors such as age, gender, accident history, address, annual mileage, car type, etc., and using a completely black-box decision-making AI in this scenario would be problematic. The customer needs motivation and an explanation for their quote and what steps they can take to achieve a better deal. FICO simply feeds different inputs to the decision-making AI over and over again to show the customer what tools they have at their disposal to influence the model's outcome. According to FICO, this approach allows the black box to become more transparent, allowing both users and those affected by the AI's decisions to understand the reasoning behind it. It is however important to notice that for AI systems deployed in industries, there are many involved stakeholders, each with different requirements when it comes to explainability. Furthermore, this approach simply is not enough for expert users working with highly technical industrial processes. The approach taken by FICO is only a partial solution, and as complexity increases and users require more data to work with, more advanced tools of explainability are needed.

2.1.1.1 Explainability Interfaces in XAI

In industrial applications, visualization mechanisms for explanations are preferred since traditional non-visual explanations are deemed insufficient. According to Kovalerchuk [19] visual and granular methods of explainability increase both the validity and interpretability of AI models, with the supporting argument being that visualizations are more appealing to human perception. Visual explanations also offer more domain-specific implementations where the design of interfaces varies depending on the specific explanatory needs of each domain. In this section, specific interface components used in a variety of real-life XAI applications will be presented.

In *Sibyl* (Fig. 2.2), the authors explore the usability challenges that come with using AI predictions for high-stakes decision-making processes [18]. They created an interface for a child welfare screening using AI predictions and found that a *Case-Specific Details* interface is the most successful option for the domain's challenges.

iForest (Fig. 2.3) is an interactive visualization tool that analyzes random forest models from different angles [20]. Users can adjust the different parameters to track changes in the predictions, and they can review the prediction process of the decision-making.

The application *MedCV* (Fig. 2.4) is a visualization system that helps practitioners make selections from medical data. The system is used by domain experts and the authors conclude that it is a tool that could prove useful for retrospectives on medical data [21].

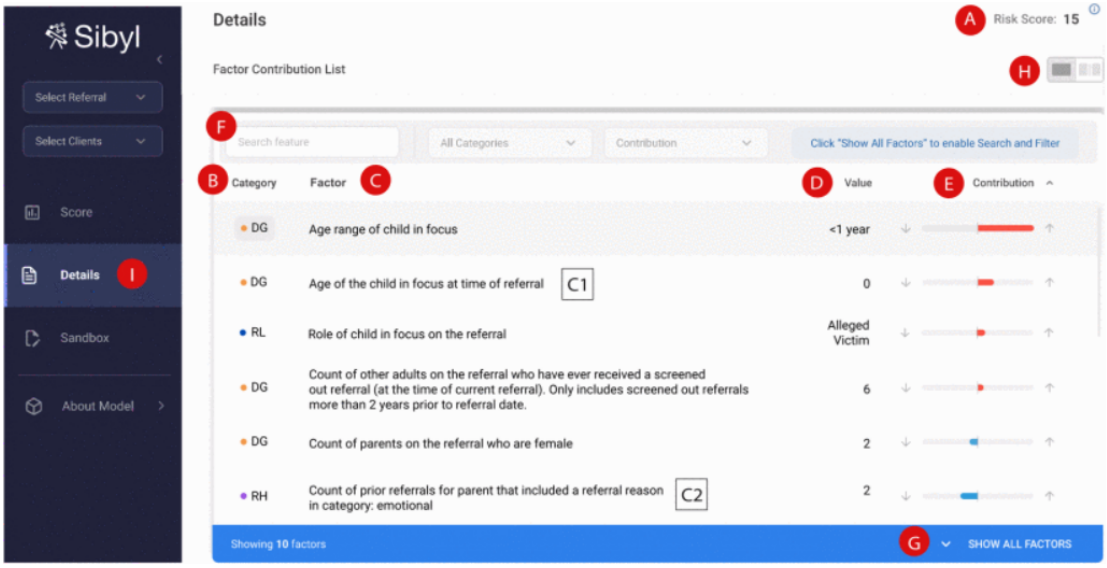


Figure 2.2: This interface displays how different factors contribute to predictions made by machine learning models about child welfare. It shows the risk score for the case, categories for each factor, a short description of each factor, the value of numeric or categorical factors, and the contribution of each factor [18].

Figure 2.3: This interface displays how iForest is used to interpret random forests with the Titanic dataset. iForest provides three different views that allow users to interpret random forests from various perspectives [20].

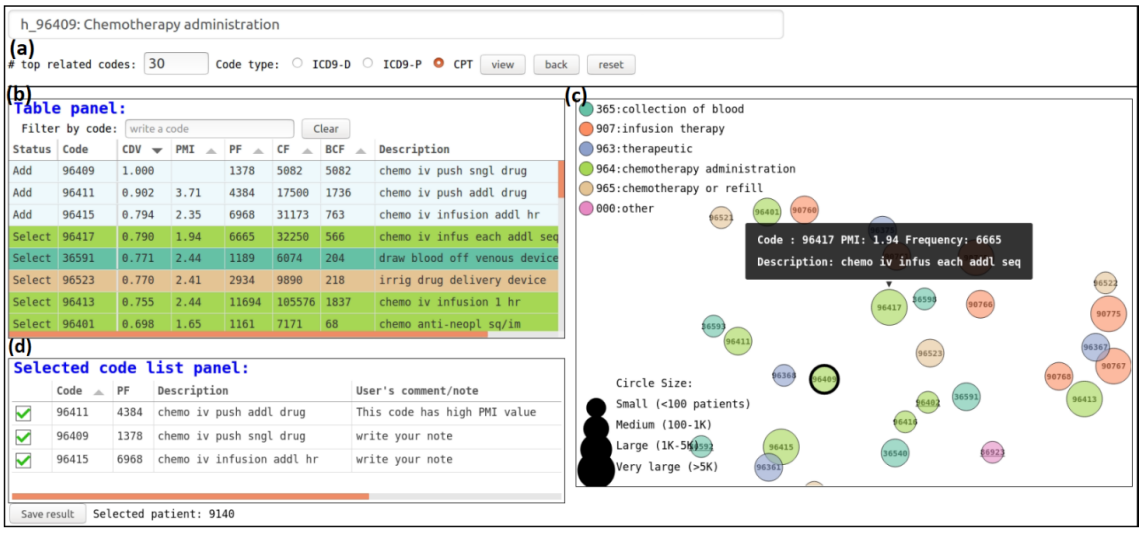


Figure 2.4: This is a screenshot of MedCV, a platform that helps physicians keep track of their credentials and continuing medical education. The screenshot shows four views that allow users to search for medical codes, view a ranked list of related medical codes, display medical code relationships in a 2D view, and keep a record of medical codes selected by the user [21].

2.2 Previous Research in the EXPLAIN Project

As a part of the EXPLAIN project [5], the Swedish partners have run several field studies to understand and elicit user requirements from the operators. The studies were conducted in 2023 and were mainly interviews. During the field studies which were conducted at two paper production plants, they investigated the workspace in which the operators work, and their thought process while solving problems, to understand them and their work process, and to identify user requirements for the XAI system. Researchers in the project have discovered that a pronounced problem with current approaches to XAI is a lack of contextual information while disregarding the task and problem-solving process of the operators. Thus, there needs to be a clear formulation of specific user requirements and contextual information that the XAI approach can adhere to. In the following section, the work process of the operators will be explained.

2.2.1 KRAFT Process and Operator Workflow

During their previous research, ABB has identified and reported on the KRAFT process in detail. This will be summarized in this section. In a round centralized control room consisting of four separate workstations placed in a circle, the operators monitor and control the KRAFT process (general description in Fig 2.5). The KRAFT process consists of four distinct phases:

- Initially, the wood chips are sorted and treated for an hour in an impregnation tank with white liquor. This prepares the wood chips for the lignin to be removed in the following step.

- After being treated, most of the lignin is removed from the wood chips. By being treated in a high-pressure digester (a large tank) at a high temperature, the infused white liquor causes the lignin to break down and the wood chips can turn into a pulp. During this phase, the *kappa* value is an essential variable for the operators, as it represents the remaining lignin after the wood chips leave the digester.
- After delignifying the wood chips, they enter a lower-pressure pipeline where they are turned into a pulp. The pulp is then washed.
- After being washed, the pulp is bleached and made ready for the following steps in the paper-making process.

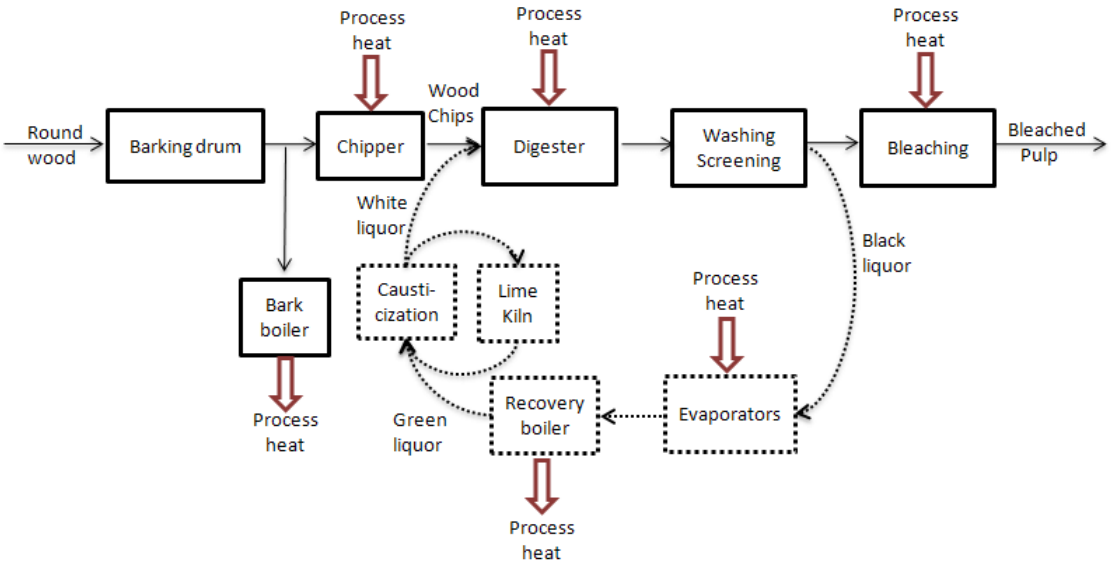


Figure 2.5: A general overview of the KRAFT process, from [94].

The stations are monitored by one to two operators close to each other which allows them to speak to one another while working. The goals of the operators are to achieve a stable and accurate kappa value for the pulp and an output that matches the tempo of the rest of the plant. Additional goals involve maintaining a high yield and alkali strength in the delignification process.

Most of the operators' work revolves around monitoring the process. Their main interaction with the system is adjusting the H-factor in the digester to achieve the desired kappa value. When there are issues in the process the operators often adjust the amount of pulp going out through the digester to match the rest of the KRAFT process. Despite the heavy monitoring work of the operators, a lot of the process is carried out with automated controllers and the operators' job is to adjust the variables of the different processes.

The tasks of the operators can be divided into three steps:

1. Chip preparation
2. Delignification
3. Washing and distribution

2.2.1.1 Chip Preparation

In the chip preparation step, the operators choose between piles of different wood chips. To do this step, they need to know what quality chips have been ordered by the customer, how much sawmill dust should be included, and the size of the different chip piles. The operators also need to monitor the impregnation process after deciding on the chips. They need to make sure that the chips move smoothly and that the amount of white liquor (a mix of hot water, sodium, hydroxide, and sodium sulfide) added into the process matches the amount of chips. When the chips enter the impregnation tank, the addition and extraction of white liquor need to be monitored to ensure that there is a balance to not oversaturate the wood chips. The temperature in the impregnation tank should be moderately high to match the temperature of the chips.

2.2.1.2 Delignification

In the delignification step, the kappa value is constantly monitored. The kappa value is the most important for the operators to keep track of, as the value has a big influence over the downstream phases of the KRAFT process. The kappa value represents the remaining lignin in the wood chips, which is a crucial variable for a high-quality product. This value is greatly affected by the quality of the incoming wood chips, such as size and humidity levels. To reduce the variance, the operators choose chips from different piles and incorporate appropriate levels of sawdust into the mixture. Because of the different quality chips, there can be deviations in the goal kappa. The operators therefore have to mentally calculate probable deviations in the kappa value. They can also discuss the change in value over the last hours with their colleagues. To understand the trends of changes in the kappa value the operators need to know the kappa in relation to the goal value, the temperature at the top of the digester, the amount of pulp the digester outputs, the moisture of the incoming chips, and the receptiveness of moisture of the incoming chips.
Further in the delignification process, the operators need to monitor the H-factor. This is the main way they control the kappa value, which is generally a stable process and therefore does not need adjustments very often. It takes multiple hours after a change in the H-factor to see an effect on the kappa value, which is another reason why the operators do not want to change it too often. The operators worry that if they enter a wrong H-factor the change will go unnoticed for multiple hours. If the operators need to change the H-factor during a working shift they often have to do some mental estimations and use trends to look at how previous H-factor changes have affected the kappa value. They therefore need to know about the deviations between the current kappa and the goal kappa, as well as the relationship between the H-factor and the kappa.
Also in the delignification process, the operators need to monitor the alkali strength. Alkali is a part of the white liquor and is crucial to break down the lignin. Daily lab results come in to complement the sensor readings on the alkali strength and how much of the alkali in the white liquor is being used. Currently, the operators do mental estimations of how much white liquor should be used based on how much pulp is currently being processed during the delignification process.

2.2.1.3 Washing and Distribution

The third step, washing and distribution, includes general monitoring of white liquor levels. Operators want to maintain a non-zero alkali level strength in the white liquor. Since the white liquor is used throughout the entire delignification process the levels of incoming and outgoing white liquor must be balanced. All of the white liquor used in the process needs to be processed in lime kiln containers and operators need to be careful to not use too much. The operators therefore need to know how much white liquor is available in the canisters, where the white liquor flows at different parts of the process, the difference between the amount of added and removed white liquor, and the status of the equipment.
Also in the washing and distribution step, the operators need to balance the output pulp based on the capabilities of the process. If there are issues further down in the process, the output pulp from the delignification process needs to be reduced. Blast tanks are used as buffers at the end of the delignification process, but the operators still need to make sure that they are not getting full as that would mean that the delignification process has to shut down which would result in the pulp reaching a too low kappa value and thus become overprocessed. The operators make mental calculations on how much the delignification should slow down so that it matches the capabilities of the rest of the process for the coming hours. The operators also discuss with other operators sitting at the other phases of the process to get an indication of the status at their respective phases.

2.2.2 Existing Challenges in the KRAFT Process

Through the field studies conducted by ABB at the plants, the researchers found several existing challenges for the operators:

- The kappa value can change greatly depending on the quality of the incoming chips; dry or wet chips retain different amounts of lignin since they are more or less receptive to moisture.
- The operator's job might be monotonously surveying the system during their entire shift if everything is running smoothly and they only need to do one adjustment.
- Because of the five to six-hour delay when doing an adjustment to the H-factor, the operators are restrictive since they do not want to risk entering a poor adjustment value.
- Since the objects and machines are co-dependent for the process to proceed, the operators need to identify and find interlocks before fixing problems.

2.2.3 ABB's XAI Solution

Via ABB's extensive research into the KRAFT process, the operators' requirements, and their workflow, ABB has identified several requirements for XAI in this specific industrial context. Through this, they have created an XAI dashboard (currently not deployed), which addresses the operator's concerns. However, as a part of the EXPLAIN project [5], IET remains an unexplored topic in this initial implementation. IET enables the operators to continuously train the system and improve its

explanatory powers and prediction accuracy. This training comes in the form of the operators actively providing feedback to the system. As a part of their research, ABB has identified several user requirements regarding this feedback. These will be further discussed in the following section.

2.3 Incremental Explanatory Training (IET)

One of the steps in the EXPLAIN project [5] is IET (see Fig. 2.1). In short, the purpose of this step is to ensure that the end-users in their respective industries contribute to the training and refining of the model. The end-user interaction serves to optimize the model by utilizing the user's feedback. Due to its novel nature, there is little research on IET and how it should be designed and used within XAI, even if it is an important step in ensuring that AI systems are successful in industrial contexts. There is, however, research suggesting that such a step is necessary [2, 14, 48], and some initial explorations as to how features similar to IET can be implemented [11, 14, 116].
Liao et al. [2] suggest that evaluations of explanations could cause users to invest more in the AI system if they know how the system will improve. Furthermore, there is research displaying promise for improving AI models based on feedback provided by users [116]. Concerning interaction modalities for providing feedback, written text feedback has previously been explored as a way to interact with AI models, with the benefit being that it is flexible and insightful [120]. However, other modes of feedback should be explored as well, which is what the research in this thesis aims to do. A prerequisite for enabling users to provide good feedback is the usage of good explanations within the XAI system [116]. Interactive explanations could benefit users by iteratively making explanations more suited to the specific domain. Explanations that are well-adjusted to the domain increase trust in the AI model [48]. Furthermore, Liao et al. [2] state that such feedback loops between the human and the AI could serve to correct outputs from the AI, which is precisely what the EXPLAIN project [5] aims to do.
Research by Teso et al. [14] proposes a framework for exploratory interactive learning, where end users can interact with and correct explanations from the AI model. Interacting with the AI and learning from it allows the user to some degree understand the reasoning behind queries and predictions done by AI models, as well as iteratively improve the accuracy of the explanations. Previous research has also shown that highlighting the relationship between user interaction and predictions influences the users' preferences for a system [117]. To achieve trust in the system it is crucial to increase transparency [28]. Concerning trust and predictive AI models, there is research suggesting that allowing users to explore contrasting features in predictions plays a role in user trust [119]. Furthermore, additional research also suggest that interactive explainable AI has the potential to improve both predictive and explanatory performance, which in turn leads to increased trust in the model [28]. Enabling such transparent interactive explanations between users and AI models as a feature of providing feedback to the system, makes the training of the system a cooperative learning process for both the models and the end users.
Another example of a similar approach is *explAIner*, a framework created by Spinner

et al. [11] which gives users the possibility to understand AI models, identify limitations with the model through XAI methods, and optimize and refine the model. In the refinement phase of the *explAIner* framework, model users can identify refinements via the XAI methods presented by the interface. Most importantly, model users can interactively identify, save, and annotate interesting findings in the model output, to further refine the model. *explAIner* aims to assist users in gaining deeper insight into the inner workings of AI models and their behavior.
With these examples as background, ABB's research has identified user requirements for the integration of IET features for the operators monitoring the KRAFT process. In this research, one goal is to confirm these and identify further requirements. The requirements identified by ABB are the following:

- The IET components need to be dynamic in required time and effort, allowing the operator to provide simpler feedback (within 10 seconds when time and effort are lower), but also more complex forms of feedback when the process is in control (30 sec - 1 min).
- IET components should not be invasive, during high-stress situations, such as process upsets where operators must prioritize attending to the process rather than giving feedback. It should be easy to back out of or pause usage of the components and not require excessive attention such as pop-ups during interface usage.
- It should be possible to provide feedback using the mouse and free text annotations or comments. Mouse-exclusive feedback is suitable for less complex feedback and free-text more fitting for more intricate answers. For free-text annotations or comments, pre-chosen alternatives should be available.
- The IET components should increase operator motivation to provide feedback, for example by communicating back how the feedback might impact the model or by other means, to assure operators that their feedback is vital. Operators are constrained in their mental models of feedback, which currently consists of direct communication with system engineers regarding maintenance and improvements of control strategies and applications. Their mental model of feedback involves being appreciated and listened to by system engineers, and eventually seeing a materialization of the input that they have provided in their system. If their feedback is disregarded, they get an explanation from the engineers as to why that is. This is a complex topic related to motivation but also has ethical considerations of how to motivate users to keep providing feedback even though it might be disregarded in the end.
- The IET components should provide a possibility to give feedback over a longer period, as some suggestions might have great initial results, but degrading results down the line (such as a few hours or days later).
- The IET components should refrain from storing personal information about who is providing the feedback.
- Operator feedback should not give direct input to re-training the model that is currently operating in the system. A data scientist should be able to access, prune, and evaluate the feedback (by comparing how it might impact the model) given by operators before it is used to retrain the model. However, there might be value for the explainability of operators in allowing them to

trigger retraining in personal, isolated models or simulations to provide a way of exploring differences in forecasts if the model is aligned more with the mental models of operators.

To this end, as specified by the research questions, this thesis investigates, 1) which types of motivations are required for sustained user engagement of the IET components, 2) which interface components enhance feedback provision, and 3) identifying design recommendations for optimizing user experience and sustaining motivation for feedback provision.

3
Theory

In this chapter, the relevant theoretical research and frameworks related to this thesis will be covered. The design approach, general theories about XAI, facilitating user motivation and engagement through UI, and the principles of UI and UX design will be described.

3.1 Double Diamond

In this research, one iteration of the Double Diamond which covers an entire design cycle was used. This methodology is well established as a model describing a typical design process while promoting divergent and convergent thinking [66]. It is widely regarded as a simple yet effective approach to design thinking. This approach has even been demonstrated to result in faster problem-solving [65]. It provides a systematic approach that balances creativity with structure. Through its emphasis on understanding during the initial phases, the Double Diamond ensures a close connection with the users' needs and requirements and the iterative nature of the process allows for continuous improvement. Designers can refine their ideas, prototypes, and solutions based on feedback and testing. This adaptability ensures that the final outcome closely aligns with user requirements. The importance of User-Centered Design (UCD) is further explored in Section 3.2. Furthermore, by discouraging rushing into solutions without thorough exploration, the Double Diamond minimizes the risk of committing to sub-optimal outcomes prematurely. However, there is a risk that too much emphasis on divergent thinking can lead to a prolonged, or endless ideation phases that does not lead to any practical outcomes. It is therefore crucial to balance creativity with feasibility and convergent solution-oriented thinking to be able to address practical issues of usage and implementation [87].
In the context of our work, this approach to the design process is very fitting as it involves close consideration of end-users. Its overall structure also provides the perfect balance of creative and solution-oriented thinking suitable for this research.

3.2 User-Centered Design (UCD)

During the design process a user-centered approach is critical as successful integration and acceptance of AI systems in industry is only possible if end users are integrated in the design process [22]. To this end, identifying concrete user requirements specific to the situation of the operators is necessary, some of which have

already been identified and described in Section 2.3 in Chapter 2. As designers it is important to completely understand the work process of the operators, what variables they work with and prioritize, what challenges they meet in their current environment, and how they problem-solve issues that arise within the production process. Integration of AI systems in already highly optimized industrial processes is especially sensitive to these factors. To address this, the aim is to gather more specific requirements from interviews with the operators, which will be further discussed in Section 4.2.2 in Chapter 4. To further understand the topic of UCD, the following sections will delve deeper into UCD with a focus on XAI.

3.2.1 Human-Centered XAI

For society to thrive and to ensure that AI lives up to its potential for progress and innovation, a human-centered approach is essential [45]. This implies a pan-disciplinary collaboration where humanistic design research plays a large role in both technological development as well as policy-making [15]. Several human-centered design practices for XAI have been highlighted in the literature, some of which are strongly related to the focus of this thesis. Below are some of these relevant practices, compiled by Auernhammer [15], with added reflections relevant to this thesis:

- **Interaction Design:** AI systems need to be designed with usable interaction capabilities in their interface to maximize usefulness. Regarding explainability, following this design practice implies a close look at how explainable interfaces communicate with the operators and what interaction capabilities they offer.
- **Human-Centered Systems:** The use of AI has an impact on organizations as well as social systems, meaning that there needs to be a focus on the humans using the applications. For example, AI systems implemented in industry processes need to be aware of and responsive towards the operators, so that they work in conjunction.
- **Participatory Design:** To ensure that the AI system is usable, users must be a part of the design and development process. This is taken into consideration in this thesis and in the previous research done by ABB, as operators of industrial processes are actively consulted throughout the development and evaluation processes. Furthermore, as discussed later in this thesis, operators were asked to participate in a participatory design workshop as a part of the research, to make sure that their needs and wishes were incorporated into to the design process.
- **Persuasive Technology:** AI systems can persuade humans into a certain behavior and in the context of explainability, it means that the system will persuade the operators to change certain parameters of the process they work with. This raises a lot of ethical considerations and practical concerns about the validity of the AI system which need to be addressed during development.
- **Need-Design Response:** Design and development of AI systems need to consider users' intellectual and emotional needs. This implies identifying user needs early on in the design process. This step is part of the *Design Thinking* process [16] and is central to the research conducted in this thesis.

To design fit-for-purpose explanations for industrial XAI applications, developers and designers require an in-depth understanding of the decision-making process of the operators, the use of explanations, and the explanation needs of the users [40, 41]. This understanding can help developers and designers create explanations that are more transparent, interpretable, and trustworthy [17]. In Section 2.3 in Chapter 2, the specific user requirements identified by ABB for the IET component, which strongly relates to the requirements of XAI in general, were discussed.

On this topic, some research suggests that it is beneficial to extract findings from social sciences to produce satisfying explanations from XAI. For example, research on what constitutes good explanations between humans could serve as a basis for designing XAI. In Miller's [95] literature review, looking at various papers from social sciences, cognitive science, and psychology, he found that explanations are contrastive as responses to counterfactuals - meaning that users do not ask why a certain event happened, but why it happened instead of some other event [95]. He also found that explanations are selected in a biased way by users. This is because users rarely expect the explanation they receive to give a complete cause of the event. Furthermore, he found that it is more effective to refer to the cause of an event rather than to refer to probabilities or statistical relationships. Lastly, Miller states that explanations are inherently social as they are a way to transfer knowledge in a conversation. When presented in a conversation they are relative to the user's beliefs about the explainee's beliefs. In our case, the explainee is the XAI system and the user might be biased when they receive information from it because of their previous biases of AI systems and the process they work with.

Furthermore, the linguist Paul Grice [96] famously compiled a set of principles describing what makes up effective communication. While communicating, people usually do so with the purpose of conveying information, and while doing so they intuitively follow a set of unspoken rules. To this end, Grice identified a set of maxims of what constitutes effective communication:

- **Quality:** The information should be of high quality and true.
- **Quantity:** Convey the appropriate amount of information.
- **Relevance:** Make sure that the conveyed information is relevant.
- **Manner:** This relates to *how* the information is conveyed, such as being clear, brief, and orderly.

Some research has argued about the relevance of these principles and criticized them for being too vague to be incorporated in computational natural language systems [97]. However, for the purpose of this research, it constitutes a good framework for examining the explanatory needs of the operators using an XAI system, especially since research highlights the connection between human explanations and XAI explanations [95].

From a designer's point of view, it is also crucial to understand how to design XAI products that satisfy the user's concerns [2, 38, 39]. Liao et al. [2] interviewed UX designers and design practitioners at IBM and created a question bank for practitioners to use when designing their XAI systems. This bank consists of typical questions that users may pose about the AI system. They assert that the effectiveness of explanations depends on the specific questions users ask, and it necessitates an understanding of user queries related to a particular AI application. The development

of this XAI question bank aims to bridge the gap between user requirements for AI explainability and the technical capabilities offered by XAI methods. Researchers in XAI often rely on an algorithm-centric view [35], which is problematic since end users oftentimes do not have deep knowledge of the technical aspects of AI. This approach leaves the users out of the loop and solely relies on ML researchers to intuit what constitutes a good explanation. One of the most common XAI approaches to explain predictions made by AI models is to list *feature importance*, which is a hierarchical list of variables influencing a prediction [36]. Although common, Liao et al. [2] argue whether such explanations are enough to satisfy users. Fully closing the gap between XAI algorithms and user needs calls for transdisciplinary collaboration and highly user-centered approaches.

Via their user-centered research, Liao et al. [2] have identified further questions that may arise for users. For example, in the case of predictive AIs, the user might want to know what minimal changes are required to produce a different prediction, and what maximum changes they can make to produce the same prediction. These are examples of *counterfactual questions* [34], which are best answered by inspection or example-based explanations. Liao et al. state that this allows the users to understand the decision boundaries of the AI prediction.

Furthermore, Liao et al. [2] clarify that the purpose of explanations is not only to make AI models more transparent but also that explanations have their own inherent utility. Through adequate user-centered explanations, the AI can expand on its prediction, and the user gains further insights into both the prediction itself, as well as the underlying topic the prediction concerns.

With this in mind, Liao et al. [2] states that it is important for AI to align itself with how humans explain things in their domain. The expert users in industrial contexts have a mental model of how their particular process and system operates. Ensuring that the AI is properly aligned with existing mental models is critical to building trust [24]. This will be further explored in Section 3.5.2.

Furthermore, limitations of human cognition are important considerations for human-computer interaction [44]. XAI methods therefore need to be aware of cognitive pitfalls such as cognitive biases. Wang et. al. [3] created a framework that would aid in reducing cognitive biases in user reasoning goals when interacting with an AI. They found that there are multiple advantages to their framework: it aids the user in generating a hypothesis, supports coherent factors when the features contradict their typical relationship, and supports access to source and situational data. These advantages show how explanations can be useful, how certain reasonings fail, and how to mitigate these failures with the help of XAI. In particular, the paper links concepts in human reasoning patterns to XAI techniques.

For example, as popularized by Kahneman in the book *Thinking, Fast and Slow* [25], there are two ways of thinking, System 1 and System 2. System 1 thinking is intuitive, fast, and effortless. The person utilizes heuristics to quickly understand situations and make decisions. This way of thinking is more prone to heuristic biases which oversimplify situations and may lead to poor decisions. System 2 on the other hand is more deliberate, slow, and analytical, where the person rationally thinks their way through situations and decisions. This way of thinking, however, is also prone to errors, such as overtrust in a wrongly calibrated AI. To avoid these biases

in thinking Wang et al. [3] state that XAI must take them into account. Again, counterfactual explanations are highlighted as an approach, where different input features lead to different outcomes or predictions by the AI. This can be utilized to avoid premature assumptions about an output, which System 1 is especially prone to. Moderating trust in the AI is also a way to avoid biases present in System 2. According to the framework, this means that the XAI should be truly transparent about its inputs and classifications of the data so that the user can detect faulty reasoning within it. With effective XAI solutions, such faulty reasonings should become clear. Once it is, the purpose of the IET components becomes evident: to provide feedback to the model explaining *why* the output is faulty.

3.2.2 User Requirements for XAI

Traditionally, XAI has primarily focused on making the naturally opaque AI models more transparent for the experts in development (e.g. [43]). Due to the rapid evolution and increasingly widespread use of AI in industry, it is important to understand the needs of explainability for the different involved stakeholders. Solely opening the "black box" of AI is not enough for true explainability [10]; what is needed is to fulfill the needs of all involved stakeholders, providing contextual explanations depending on their requirements [42]. In this context, Langer et al. [1] clarifies five different stakeholders: regulators, deployers, developers, users, and the affected parties. Each stakeholder group has different topics of interest which need to be satisfied using explanations that clarify different aspects for each group. Langer et al. [1] focus their research on *how* these specific stakeholders' interests can be satisfied. According to Langer et al. some specific topics of interest for the users in particular are: trust, controllability, debugability, confidence, effectiveness, satisfaction, safety, and the ability to gain scientific insights from the system. XAI systems in industry should ideally provide explanations that satisfy these concerns to ensure that they are useful to the end-users. Furthermore, these are important considerations to our research as the IET components could enable the end-users to give feedback on these topics. XAIs need to be controllable and debugable to ensure their usefulness and without any method of doing so, these AI systems will fall short of their promises and end up being unsuccessful. In turn, AI systems that can be incrementally trained, controlled, and debugged, increase trust in the system. The IET components are attempts to satisfy these concerns through iterative training with the end-users.

3.2.3 XAI Evaluation Methods

When it comes to XAI, it is not possible to quantify criteria such as accuracy or performance [30], criteria that are typically used to measure the performance of AI models in general. There are however some metrics for evaluating XAI systems which include the goodness of explanations, user satisfaction, understanding, curiosity, motivation, trust, and performance [29]. A commonly used criterion for XAI is interpretability [31], checking if the system can explain its reasoning, and through that developers can verify whether the reasoning is correct concerning the other

criteria such as accuracy. Doshi-Velez and Kim [13] write that there is still little knowledge of what interpretability is and how it should be evaluated. Looking into the research of evaluating XAI systems is useful to identify some of the most common metrics for measuring their success. Since the goal of the IET components is both to continuously improve the accuracy of predictions and the explainability of the model, insights from commonly used XAI evaluation methods could serve as useful input to the IET components design.

3.2.4 Participatory Design

A general trend in design is the movement towards an increasingly close collaboration with end-users. Historically, UCD has been one of the initial advancements in this direction, and in recent years, participatory design has gained momentum as a viable method. Solely utilizing a UCD approach is argued to not completely address the scale and complexity of modern-day systems [98]. It could be argued that effective design approaches should rise above the creativity of a single stakeholder and be open to multiple angles of influence. This is addressed by participatory design. In turn, a growing notion within participatory design is co-design [98]. Instead of researching the users, the users become partners in the design process and actively contribute with their own ideas and experience. Incorporating participatory design or co-design into design practice changes what is designed, how it is done, and who designs it. It is especially useful early on in a process to establish new ways of thinking and formulating problems [99]. During participatory design and co-design, the roles of the users change from being passive subjects to having an influential role in gaining insights and developing ideas. The designer simply supports the user in this process by providing expressive tools to realize the user's ideas. Since the end-users in this particular research are highly specialized experts within their domain, participatory design is a complementary approach to ensure the designs' usefulness.

3.3 Facilitating Motivation and User Engagement

Since one of the primary goals of this research is to discover what interface components contribute to motivation to provide feedback, the science behind motivation in relation to UI and UX design was an overarching topic that guided decisions.
Motivation itself can be divided into two parts: *intrinsic* and *extrinsic* motivation [23]. *Intrinsic* (or internal) motivation is a motivation that comes from within; the user does a certain activity because they enjoy it, not because they are forced to. This is often curiosity-driven, meaning that if the user is curious about the system they are more likely to use it. *Extrinsic* (or external) motivation on the other hand is a motivation that comes from the user expecting some kind of consequence or outcome, such as a child doing their homework to avoid negative consequences and not because they enjoy doing their homework [23]. In the context of this research, *intrinsic* motivation is how to make the act of giving feedback to the system valuable to the operators without a direct positive consequence. *Extrinsic* motivation, however, is how to make the act of giving feedback to the system motivating in itself. Concerning user engagement, Darejeh et al. [53] has identified some common issues

with software which has a negative impact on it. Three of these are especially relevant in the context of this research:

- **Unpopular systems:** Generating content into software is oftentimes essential for its success. However, if the system itself is considered unpopular by users, motivation to do so will stagnate.
- **Difficult systems:** For many people, learning new software is a difficult undertaking, and if it is too difficult, users will stop engaging with the system.
- **Usage of real identities:** In many cases, users wish to avoid using their real identities, as they might experience uncertainty about the correctness of their answers, or they might simply feel shy to express their opinions. This is also a requirement identified by previous research done by the Swedish partners in the EXPLAIN project [5] (listed in Section 2.3 in Chapter 2).

Based on these common pitfalls to user engagement and the research into motivation, it is important to consider them during our design process. Furthermore, there are additional techniques to increase user engagement.

3.3.1 Fogg Behavior Model

For promoting user behavior, the Fogg Behavior Model is useful [103, 104]. The model specifies that three things need to converge in order for a behavior to happen: motivation, ability, and a trigger. As a condition for taking action, motivation and ability must reach threshold levels for users to take action. Furthermore, the users must be triggered to perform the behavior, as is visualized in Fig. 3.1. Fogg also specifies that small easy-to-do steps build upon each other, and without much resistance, the user eventually engages in more difficult tasks on their own [103]. The Fogg Behavior Model outlines important considerations for the IET components, especially in regard to motivation. Looking closely at what contributes to motivation is important for discerning how to design to support it. Furthermore, since the goal of this research is to ensure prolonged engagement in the behavior of providing feedback, it is important to consider both ability and triggers as well. The following section will delve deeper into these components within the Fogg Behavior Model.

3.3.1.1 Motivation

Concerning motivation, Fogg divides it into three main components: sensation, anticipation, and belonging [105].

Sensation implies that users are motivated to perform certain behaviors if they lead to pleasure or avoidance of pain. To increase this type of motivation one suggestion is the utilization of gamification elements, this is further discussed in Section 3.3.2.

Anticipation is divided into two parts, hope and fear. Users are motivated to certain behaviors in accordance with their hope of something good happening, or their fear that something bad will happen. Increasing this type of motivation could also relate to gamification elements, such as the progression of narratives, where engaging in a certain behavior progresses a story.

Lastly, belonging refers to the social nature of humans, and that we pursue certain behaviors with the goal of improving our social status and sense of belonging. Similarly to the last two types of motivation, belonging can also be increased via

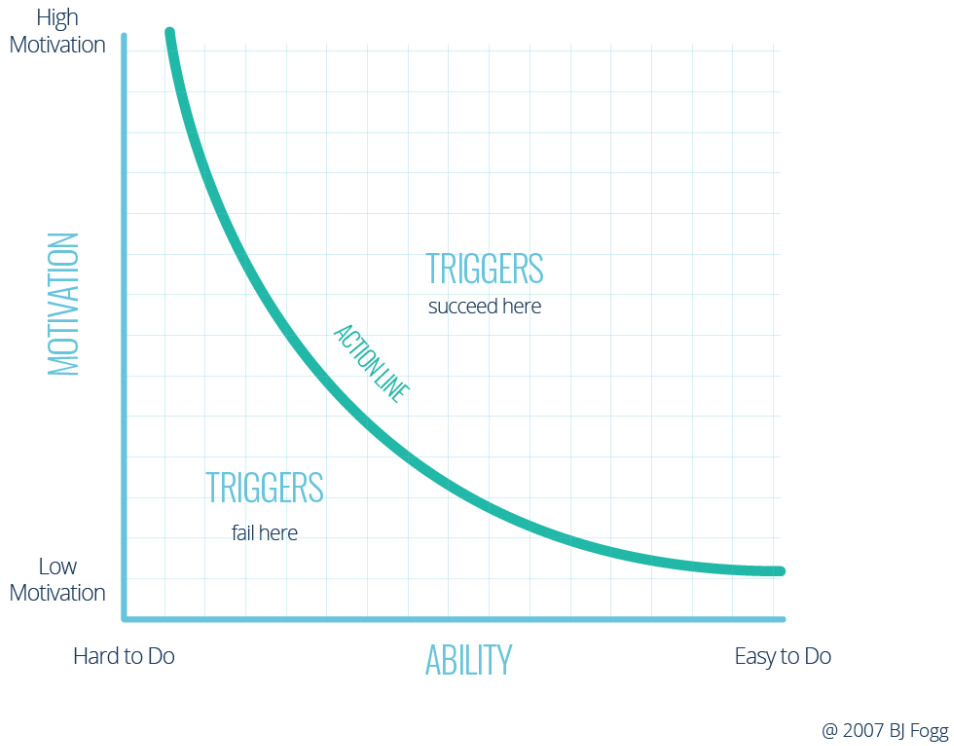


Figure 3.1: Visualization of the Fogg Behavior Model from [105].

gamification elements such as leaderboards, where the users can see their contribution towards a shared goal. This could also motivate users to share their knowledge. Gamification will be further discussed in Section 3.3.2.

3.3.1.2 Ability

Ability can be divided into six distinct components. The ability to perform a behavior depends on these factors:

- **Time:** The less time it takes to perform an action, the more likely the behavior is to happen.
- **Money:** Cheap behaviors require less motivation.
- **Mental Effort:** To ensure a behavior is repeated, it should not require unnecessary cognitive effort.
- **Physical Effort:** Physically demanding behaviors require more motivation.
- **Social Deviance:** If an action involves deviation from social norms, people are less likely to do it.
- **Non-Routine:** Behaviors that easily fit into daily routines are easier to partake in.

Many of these considerations, such as mental effort and time, overlap with research in UI and UX design (see Sections 3.4 and 3.5). Put together, these considerations will provide necessary insights into designing the IET components.

3.3.1.3 Triggers

Regarding triggers, Fogg refers to external factors that contribute to facilitating a behavior. Fogg clarifies three main categories of triggers:

- **Sparks:** These are triggers suitable for users with low motivation but with high ability. Their purpose is therefore to increase motivation.
- **Facilitators:** These are triggers suitable for users with high motivation but low ability. Their purpose is to increase the ability to perform an action by simplifying its initiation.
- **Signals:** These triggers are for when both motivation and ability are high. Oftentimes these are simply notifications or reminders to engage in an action.

To ensure that the IET components facilitates sustained usage over time, explorations of how these triggers can be implemented in the design are beneficial.

3.3.2 Gamified Approach Towards Engagement

In this section, some techniques relating to gamification will be highlighted. One very common approach to designing interfaces that increase user engagement is gamification. In short, gamification is the usage and implementation of video-game mechanics into non-game-related software [53]. This is an emerging area of research within Human-Computer Interaction (HCI) [54], as it promises to increase positive behaviors such as motivation and engagement in usage. Overall, a systematic review by Darejeh et al. [53] on gamification has demonstrated that 81% of gamified software has positive results on user engagement and motivation.

3.3.2.1 Common Game Mechanics in Gamification

According to a systematic review on gamification by Darejeh et al. [53], some of the most common game mechanics implemented in software are: stories, a clear goal, challenge, time limit, progression, immediate feedback, and rewards. According to Xu [56], rewards are the main motivator for gamified systems. Rewards are usually triggered by a reward trigger, which is a predefined action that leads to a reward upon completion. The number of rewards depends on the number of triggers. For example, users can be rewarded for inputting certain data or answering questions [55]. There are also different reward types, where the most common ones are: points, virtual money, levels, badges, status, and achievements. These rewards can be used in three different ways, achievement game, in-game, and out-game. Achievement game rewards are simply a sign of progression once a task is completed. In-game rewards are used within the software to upgrade certain things or reach the next level. Out-game rewards are rewards that relate to things the user can get outside the software environment, for example, food discounts or bonuses [55]. Furthermore, these game mechanics can be incorporated into the UI, represented as gamified graphical themes such as fantasy or Sci-Fi. However, the degree to which such graphical elements are added depends on the nature of the software and its use case [53]. In certain situations, it is simply not appropriate or needed.

3.3.2.2 Gamification in Industry 4.0

Some research has been done regarding gamification in relation to Industry 4.0 and has found that it can be used as a tool to increase *intrinsic* motivation and to successfully introduce new technologies [57]. Some rewards, such as the out-game rewards described earlier, contribute only to motivation in the short term and are not capable of creating *intrinsic* motivation [58]. One suggestion as to why gamification is superior in motivating employees in the context of industry is the establishment of clear goals and purposes for the tasks in question [59]. There are however challenges concerning gamification in industrial contexts, such as whether it is the right tool for proper alignment with the goals of the industrial process [60]. Furthermore, it is important to design gamified software together with the users, as the techniques would fall short if they are applied against the interests of the users [61]. Again, this emphasizes the importance of UCD practices.
Some important characteristics of gamification implemented in industry have been identified in the literature. Such as the establishment of clear goals, incremental goals, progression mechanisms, and statuses, which keeps the users engaged [62]. To mitigate the negative impact on *intrinsic* motivation, it is also recommended to avoid emphasizing scoring systems, as it could become a primary focus for the users [57]. Perhaps most importantly, gamification processes should not interfere substantially with the user's workflow [57, 61]. Furthermore, the usage of these techniques should be voluntary, as forcing these features upon users could have negative effects on acceptance [63]. The gamification techniques used should therefore be subtly implemented and voluntary, to achieve increased *intrinsic* motivation while not disrupting the workflow of the users.

3.4 User Interface (UI) Design

Since one of the goals of this research is to prototype a UI, the design choices are motivated through established literature and used so that the chosen approach fits the context. To this end, Cooper et al.'s widely regarded book *About Face* [49] is used as a reference. In this section, some relevant aspects of the book will be covered.

3.4.1 Incorporating Expert Users Into the Design Process

For complex domains, domains that have a lot of technical difficulty, or domains with a lot of legal considerations it is important to meet with expert users. Cooper et al. state that expert users can provide many important insights into the product [50]. However, Cooper et al. also state that while the users are knowledgeable, they are not designers, meaning that designers should look to the expert users for problems they are facing rather than solutions. Cooper et al. also highlight the importance of receiving guidance from expert users throughout the design process. For this research, it is crucial to keep a continuous dialog with the expert users, i.e. the operators, to get their insights on the design of the interface while also checking with them that the delignification process and that their needs have been understood correctly. Furthermore, to ensure a user-centered approach, it is important

to understand what type of interface the operators wish to use. In the following section, the different natures of UIs will be discussed.

3.4.2 Interface Posture

According to Cooper et al., a product's posture is the way it presents itself to users and different hardware modes such as desktop computers or mobile phones affect the posture [51]. For desktop applications three types of posture are identified: sovereign, transient, and daemonic. In the industrial setting in which this research is conducted, the operators work on desktop computers with keyboards and mice. The application the operators use has a sovereign posture, meaning that it is an application that takes all of their attention. Other qualities of sovereign-posture applications are that they have many functions and features that run continuously and take up the entire screen. Cooper et al. also write that sovereign-posture applications allow users to work in a state of flow due to the size of these applications. Furthermore, users of sovereign-posture applications spend little time as beginners relative to the total amount of time they will eventually spend on the application. Some design principles for sovereign applications, as identified by Cooper et al., are:

- Optimize for full-screen use
- Use minimal visual style
- Provide rich visual feedback and support rich input

Transient posture, on the other hand, are applications that are invoked when needed and disappear quickly so that the user can continue with their work which is usually done on a sovereign application [51]. Cooper et al. write that because of the temporary nature, users do not become very familiar with transient applications. Because of this, the application's interface should be simple, helpful, and obvious so the user does not get confused. Some design principles for transient applications, as identified by Cooper et al., are:

- Applications must be to the point, clear, and simple
- They should be limited to a single view, in a single window
- They should be able to launch their previous position and configuration

Cooper et al. recommend that transient applications have bolder graphics to help the users orient themselves when the application starts. They should also have instructions built-in, for example with clearly labeled buttons. Furthermore, the application should have direct and explicit feedback to the user. To keep the application simple it is important to keep all the important information – and thus the user's attention – to a single window. The choices made by the user in the previous session will most likely hold true for the next session, and it is therefore good to give the application a memory to allow it to open in the same shape it was left in. These guidelines are important as they guide how the IET components should be designed concerning posture.

3.4.3 Avoiding Excise

Excise tasks are tasks the user completes but that do not go towards completing their goals [52]. There are many types of excise for digital products and applications. One

is navigational excise, where the user has to navigate through too many unnecessary or difficult steps to reach their goals. Another type is skeuomorphic excise where mechanical representations of elements in the digital environment do not fit or work with their digital counterparts. It is also important that the skeuomorphic elements do not take up too much space on the screen, particularly in sovereign applications. Modal excise is a form of excise where interruptions in the form of dialog windows, error messages, confirmation messages, and notifications disrupt the user's flow of working with the system. Lastly, Cooper et al. write about visual excise, where the visual style of an application is exaggerated and hinders usability. Cooper et al. also point out that what one person sees as excise might be what another person sees as a goal-directed task. The matter of excise is therefore highly contextual and depends on the user's goals. The amount of excise may also vary depending on the posture of the application. Cooper et al. suggest the following to eliminate excise:

- Reduce the number of places a user must navigate.
- Provide signposts since users navigate with persistent objects such as application windows.
- Map controls to functions to reduce cognitive load.
- Avoid hierarchies since the abstract notion of a hierarchy is difficult for users to grasp.
- Avoid replicating mechanical models to eliminate the skeuomorphic excise.

In the context of the IET components, it is important to have as little excise as possible to not take the operators' focus away from the important tasks at hand.

3.5 User Experience (UX) Design

The overarching goal of this research is to create a good UX for the operators using the IET components. Through the approaches mentioned (UCD, Human-Centered XAI, techniques for facilitating motivation and user engagement, and UI design) the goal is to achieve a convergent design adhering to all these aspects. To achieve this goal, it is important to also cover some of the basics of UX design. In the field of UX research, the Nielsen Norman Group [76] are world-leading experts. Below is a selection of some important considerations when it comes to UX according to Nielsen Norman Group:

- Design for users, not yourself
- Simple designs are better designs
- Simple interactions for key features
- Design for how users behave
- Design decisions depend on the context

A general trend in this list is the demand for UCD, created from and catered to the needs of the users. Without a fundamental understanding of the users and their context, realized in the design, the UX will fall short despite adequate functionality.

3.5.1 Interaction Cost

Regarding simplicity, research has highlighted Interaction Cost as an important consideration [77, 78]. In short, interaction cost refers to the cognitive effort present

in a design, similar to excise as defined by Cooper et al. [52]. Ideally, the information required by the users should be present without performing any actions. This, however, is rarely the case and users need to navigate, type, switch attention from one component to another, and remember information to enter at another part of the UI. All those actions require cognitive effort, which should be considered. Some research also suggests interaction cost should be modeled in adaptive interfaces, to keep track of how users perceive the interface [79]. In the context of the IET components, the operators might need to remember why a certain prediction was good or bad, or input information from other parts of the interface. Presumably, the IET components should therefore minimize the need for the operators to remember all this information by themselves and should allow the operators to view previous predictions and their corresponding feedback, as well as simplify how information is fed into the IET components.

3.5.2 Mental Models

Research also highlights the importance of considering mental models, as it is one of the most important concepts in HCI research [80, 82]. Furthermore, mental models are also explored in the context of XAI systems, to understand the degree of the end-users' understanding of AI systems [81]. Mental models are compiled of what users believe about a system, how it works, what it does, and how it does it, which has a strong influence on how they interact with the interface [83]. For designers, it is important to clearly communicate the nature of the system, so that the users can form an accurate mental model of it. Users form mental models based on the world around them, for example, from previous similar interfaces they have used or are accustomed to. In the context of the IET components, it is important to consider two things: the mental model the operators have of the process they work with, and clearly communicate the nature of the IET components to them. If the IET components do not fit into this already established mental model, or if it fails to communicate what type of system it is, the design could fall short of their expectations and could end up not being used.
All the theories discussed in this chapter were used throughout the process of this thesis. How these concepts, guidelines, and frameworks were utilized will be further discussed in the following chapter.

4

Methodology

In this chapter, the methods used in this research will be highlighted. The overarching structure follows the double diamond, which will be described. For each phase of the double diamond, corresponding complementary methods that guided the research to completion have been chosen. This chapter outlines and describes the overall approach, as well as each method in detail.

4.1 Double Diamond

The double diamond is split into four distinct phases as described by Kochanowska et al. [66]: discover (diverge), define (converge), develop (diverge), and deliver (converge). Thus, the work was carried out according to these phases. The first phase, **Discover**, consists of broadening the problem horizon, and thoroughly understanding the problem space, context, and users. In the context of this research, this means a deep dive into the working conditions, challenges, and problem-solving methods of the operators. Through this stage, the aim was to understand the needs of the operators concerning XAI and IET components (see List of Acronyms on page ix). This laid the foundation for the following phase in the double diamond. The second phase, **Define**, involves solidifying the data gathered in the previous phase. This is where the design challenge is framed and an understanding of the actual design problem emerges. The third phase, **Develop**, is where the design work begins. Through insights gained in the previous phases, new solutions meeting the users' needs will be developed. In the context of this thesis, a user interface for providing IET was created. Finally, the fourth phase, **Deliver**, involves iterating on and refining the design. In the context of this thesis, this phase consisted of an evaluation of the user interface and further refinement of the design. This stage concluded with a thorough description of the final design. An illustration of the double diamond and the phases that were carried out can be seen in Fig. 4.1.

4.2 Discover

During this phase, the main focus was on analyzing and summarizing related literature, with the purpose of thoroughly understanding the related domains. Expert interviews as a method are also highlighted to gain deep insight into the area.

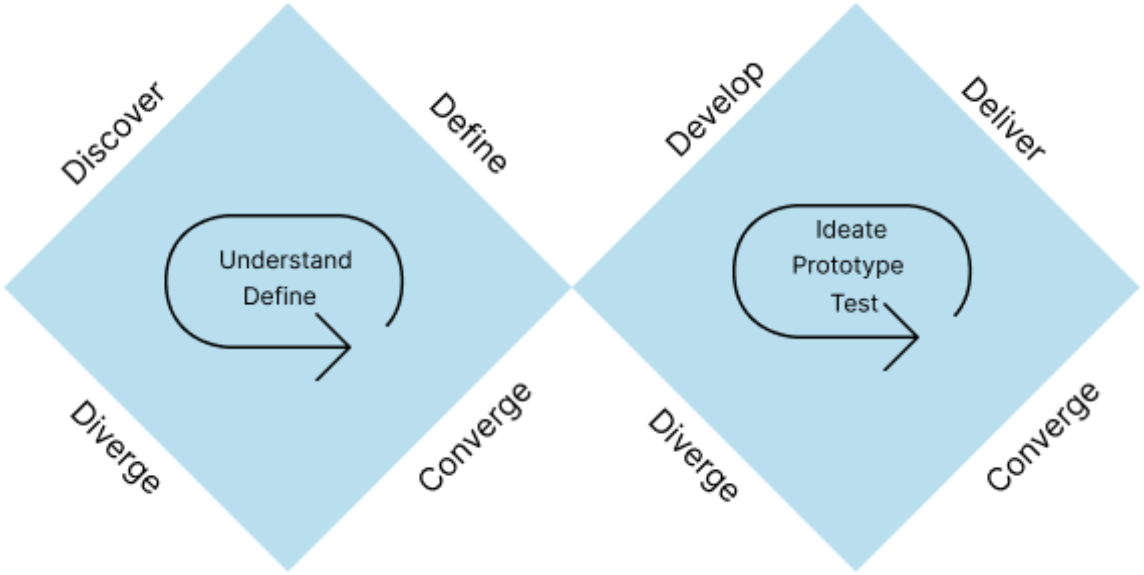


Figure 4.1: Illustration of the double diamond.

4.2.1 Background Research

As part of the background research, literature was analyzed to find what was relevant to this project. Our main focus was on XAI, industrial processes, and UX/UI design. By finding and summarizing the important literature a general overview of the research already done in this area, along with identifying research gaps, contextualizing the research became possible [67]. Understanding the current state of knowledge is crucial for developing a robust theoretical framework and research methodology. Both the industrial process itself and the operators' workflow are of great importance, as implemented XAI cannot be successful without a deep understanding of both. This initial background research deepened our insights, gave us a basic understanding, and offered the necessary background for us to further build upon with our own research. A summary of the analysis can be read in Chapters 2 and 3. This step led the way for conducting expert interviews.

4.2.2 Expert Interviews

Expert interviews are unique, compared to observational studies or surveys, as they offer researchers great insight during the exploratory phase of a project [69]. The otherwise time-consuming process of collecting data can be shortened with expert interviews, as the experts have unique insight into the systems they work with. Furthermore, both the researcher and the interviewee share a common interest in the field, which could serve to enhance the participant's motivation to engage in the interview. For the purpose of this thesis, expert interviews are crucial, as the KRAFT process and the operator's work are unfamiliar domains to us.
Not only are the expert interviews important data-gathering methods, but observational data collected during the interviews was of great value to the design of the IET components. As stated by Beyers et al. [68], observational data is beneficial as a complementary method, as the data remains pure the interview subjects cannot

unintentionally misrepresent it. However, it cannot be solely relied on.
One of the common sources of bias in expert interviewing is relying too heavily on the participant's memory. To address this, Beyers et al. [68] suggest avoiding open-ended questions which are more demanding in that respect, and thoroughly preparing for interviews by researching the area of interest beforehand. Another important consideration is the fact that the participants are experts within their domain, and not within the domain of XAI, which is what the interviews are about. Although some participants do have previous experience with XAI systems, researched as a part of the EXPLAIN project [5], the questions had to be framed in a familiar context for the operators. To address this, the interview questions were based on research connecting XAI to social and cognitive sciences [95]. In Section 3.2.1 in Chapter 3, this connection was highlighted and the interview questions are framed as if the operators were communicating with a colleague, rather than an AI. This served two purposes: grounding the questions in a familiar context, while still being relevant for our research purpose. How this was done will be further discussed in Chapter 5.

4.3 Define

This phase is primarily focused on framing the design challenge and distilling the previously collected data into usable knowledge. Here, the thematic analysis as the chosen method is highlighted.

4.3.1 Thematic Analysis

The interviews were analyzed through a thematic analysis as part of the *Define* phase of the double diamond. For the coding and construction of themes, the software MAXQDA 2020 [100] was used. Through the analysis, the aim was to distill the interview data into further user requirements and insights for the IET components. Thematic analysis is a method used to analyze qualitative data by assigning it with themes and codes [101]. The general process for a thematic analysis involves transcribing interview data, extracting codes and subcodes that summarize the transcription, and placing these codes into themes representing the general concepts. Thematic analysis can be divided into three subcategories: inductive, deductive, and reflexive. For the purposes of this research, both deductive (creating themes and codes before the interviews through the interview questions) and inductive (creating themes and codes from the findings of the interviews) coding were used. Using both types of coding means that a base can be established through deductive coding while remaining open to new themes and codes that may emerge during the interviews. Relying solely on deductive coding would significantly restrict the variety of themes and codes derived from the interviews, contradicting the semi-structured nature of the interviews.
According to Braun & Clarke [101], a thematic analysis consists of six distinct steps. Step one consists of familiarizing oneself with the data. This involves transcribing, reading, and re-reading the data so that the gist of each interview is grasped. Step two consists of generating initial codes so that interesting findings in the data stand

out. Each code represents several related instances of citations from the transcription. Step three consists of searching for themes and collecting related codes into them. All relevant codes get grouped into these initial themes. Step four consists of reviewing these themes and double-checking if the codes fit into them, and if the themes themselves represent the dataset. Step five consists of specifying the names of each theme in relation to the coded segments, and to understand on a more holistic level what the themes therefore communicate. Finally, step six consists of producing the report, which is a final analysis where the most compelling data points can represent each theme and sub-theme. The process of doing thematic analysis can be found in Chapter 5, and the resulting themes and sub-themes can be found in Chapter 6.

The results of the thematic analysis were used to guide the next step of the design process, *develop*, where the interface for the IET components were based on the insights and requirements identified in the interviews as well as the previous research done in the EXPLAIN project [5].

4.4 Develop

The *develop* phase is focused on utilizing methods with the purpose of establishing the design space and prototyping design solutions. The methods need to provide insight and direction for the design challenge, while still being sensitive to the influence of the end-users. A set of methods were therefore chosen for this purpose.

4.4.1 Brainstorming

Brainstorming was the initial step in the design iterations. According to previous research, brainstorming is an effective method for increasing creativity and helps produce large quantities of ideas addressing real-life problems, ruling out criticism, and the combination of ideas [70]. In the context of this research, it was beneficial to use graphic methods of brainstorming. This is because the result of this study was to produce an interface for the IET components and thus, conducting graphic methods of brainstorming allows us to generate ideas for the interface. To this end, *brainwriting 6-3-5* was used to create suggestions for the design of the IET components.

4.4.1.1 Brainwriting 6-3-5

Brainwriting 6-3-5 is a common brainstorming technique [73]. It involves six people identifying a problem statement describing the problem they are aiming to solve, writing this statement at the top of a piece of paper, and then producing three ideas addressing the problem in five minutes (hence 6-3-5). After five minutes have passed, the participants pass their paper to the person on the right and the process is repeated until the paper is filled. This is followed by a group inspection and discussion of the ideas, where some are selected for further development.

4.4.2 Participatory Design Workshops

Participatory design is a method of producing artifacts and systems while utilizing the knowledge of people who work with technologies [111]. Participatory design is about designers, researchers, and users creating solutions together. Spinuzzi [111] explains that a lot of the knowledge created from participatory design is tacit, meaning that it is implicit and holistic; it is the things that people know about but do not put into words. They state that most participatory design research consists of three stages, 1) Initial exploration of work, 2) Discovery process, and 3) Prototyping. In this research the three stages were utilized, where the first stage was done with the operators, and the second and third stages were done in collaboration with UX designers and operators. Taking this approach of participatory design means that together with the participants of the workshops, new knowledge about what the users want and need can be discovered, and the participants' ideas can be used as a basis for the interface being created.
Participatory design aims to enable the users to be the experts of their own experience. This shift aims to alter the roles of the users, researchers, and designers, moving away from the typical roles of UCD [98]. For this type of session to be successful the users must get the appropriate tools to be able to express their experiences, needs, and creative solutions. The researcher, who may also be a designer, becomes a facilitator instead of serving as the translator between the users' needs and the design. As a facilitator in the session, the researcher leads and guides the participants, as well as provides scaffolds to support user's creative expressions and provides a clean slate for those completely new to the process. When the users become participatory designers in the process, it might seem that the designer has nothing left to do, but their skills are still required, especially when the scope becomes larger and more complex. The designers also provide expert knowledge in the session that the other participants do not have [98].

4.4.3 Affinity Diagramming

Affinity diagramming serves the purpose of meaningfully externalizing and clustering ideas [112, 114]. A benefit of affinity diagramming is that it allows observational insights to be recorded and analyzed. In short, this method aims to synthesize and capture insights, concerns, or opportunities of a design on Post-it notes. These notes are then clustered based on affinity, meaning that thematically similar notes get grouped together. These groups then create themes, similar to a thematic analysis.

4.4.4 Prototyping

A rapid prototyping process as part of the *develop* stage of the double diamond was chosen. A rapid prototyping process usually consists of fast-paced iterative work open to user involvement. This is a suitable approach to the *develop* stage, especially in conjunction with a participatory design session.
A rapid prototyping process includes creating a basic prototype (oftentimes a paper prototype), followed by a mock-up that has limited functionality, and finally a semi-functional prototype [72]. Wilson [71] states that the completeness of a UI prototype

depends largely on the goals of the designer and the purpose of the prototype. If the purpose of the prototype is to serve as a tool for communication, it only needs a part of the functionality implemented. If the purpose however is to use the prototype to conduct testing to enable further iterations of the design, it should have a higher fidelity and more functionality in place [71]. Some benefits to UI prototyping, according to Wilson [71], are:

- They provide a way to test questions related to the specific products that cannot be answered by generic guidelines and research.
- Are a way to evaluate a UI concept in a tangible way.
- Allows for addressing meaningful feedback from users.
- Improves the completeness and quality of a product.

4.5 Deliver

The *deliver* phase is focused on refining and performing final iterations on the developed prototypes. To this end, some methods for evaluating the design to highlight its strengths and weaknesses, have been chosen.

4.5.1 User Experience Evaluation Methods

There are many ways to evaluate interfaces in UX research [90]. Thus, to evaluate the design multiple evaluation methods were used. As a part of the *deliver* phase of the double diamond, the methods described in this section should provide insights into the design itself and further guidance for future iterations.
Evaluating UX is challenging since there is a need to tap into the user's inner reasoning and not only quantitative data such as task completion [90]. A systematic review on the subject by Rivero & Conte [90] states there is a need for methods that allow for subjective qualitative data that assist the evaluators in understanding the cause of the user's experience. Quantitative data is also an important consideration specifically in industrial contexts, as the design has to adhere to the industry demands. To this end, some evaluation methods suitable for both sides of the spectrum are highlighted.

4.5.1.1 A/B Testing

Testing is crucial to ensure that the software and user experience of a product meet the requirements set by the stakeholders. There are many different ways to test software, A/B testing being one of them. A/B testing is an evaluation method used to compare two different types of software in a certain context. It is done by splitting a group of users into two or more groups and having them test different versions of the software [88]. The order in which the two designs are presented to the groups was changed to avoid bias. Multivariate testing is a form of A/B testing, but where A/B testing presents users with two completely different options, multivariate testing lets users test several options that have more or less variation between them [75]. In this research A/B testing was conducted, as there was enough variance of the IET components between the two prototype versions for A/B testing to be a suitable

method. A/B testing can also be useful for the triangulation of data, meaning that it is used as one evaluation method together with other methods to ensure the results are accurate. This is because A/B testing by itself does not give enough insights into the user's attitudes and it does not give the designer space to ask further questions about the user's behavior [89]. Taking these insights about A/B testing into account when conducting evaluations for the IET components was important. To address this, this evaluation method was used together with the think-aloud protocol and summative interviews, which will be described in the following sections.

4.5.1.2 Think-Aloud

Think-aloud is an evaluation method commonly used within UX. It is a method where the user uses a system while speaking out loud, narrating what they are doing and their thought process behind it. This allows the designers to understand the cognitive process of the user, as well as their mental behavior and general insights into their thinking when using the system [92]. The benefits of think-aloud are that you find out what the participants think of the design presented to them while also seeing what their misunderstandings of the system are. When users misunderstand a part of the system the designer needs to change it and through the think-aloud protocol the designer usually also finds out *why* the user misunderstood a part of the design [84]. For this research, think-aloud is a good evaluation method as it lets the operators test the system in an environment that is natural to them and that better represents the real-world situation in which they would use the IET components. Some suggestions for improving think-aloud sessions, based on Fan et al.'s [93] research, are:

- Conduct a practice session before the actual think-aloud sessions to help the user get used to verbalizing their thoughts and actions.
- Use neutral language to not prompt the user into giving certain types of answers.
- Keep the interaction between the designer and the participant to a minimum.
- Pay attention to the user's actions, what they verbalize, as well as *how* they verbalize it.

The think-aloud technique is also beneficial in that it enables a great variety of qualitative analysis methods [115]. Depending on the research question, researchers can use both deductive and inductive coding frameworks to categorize the data and identify common patterns and themes. Even quantitative methods could be used which focuses on the frequency of responses, patterns, and distributions within the data [115]. In general, the think-aloud method is beneficial for understanding the inner reasoning of the users while they are interacting with an interface. To this end, this method was part of our evaluation of the IET components, used to reveal and understand any usability issues.

4.5.1.3 Summative Interviews

Finally, interviews were chosen as a method for the evaluation. Interviews are an increasingly common method used for evaluating UX [127]. As a part of an evaluation, interviews shed light on broader aspects of the design. It allows the

user to reflect on its implications. Coupled with the other evaluation methods, this should result in an effective triangulation of the data. As with most interview data, a thematic analysis is a suitable method of analysis [101].

5
Execution and Process

In this chapter, the utilization and execution of each method described in Chapter 4 will be presented. The procedure followed the double diamond method as was described in Section 4.1 in Chapter 4, consisting of the phases Discover, Define, Develop, and Deliver, as shown in Fig. 5.1.

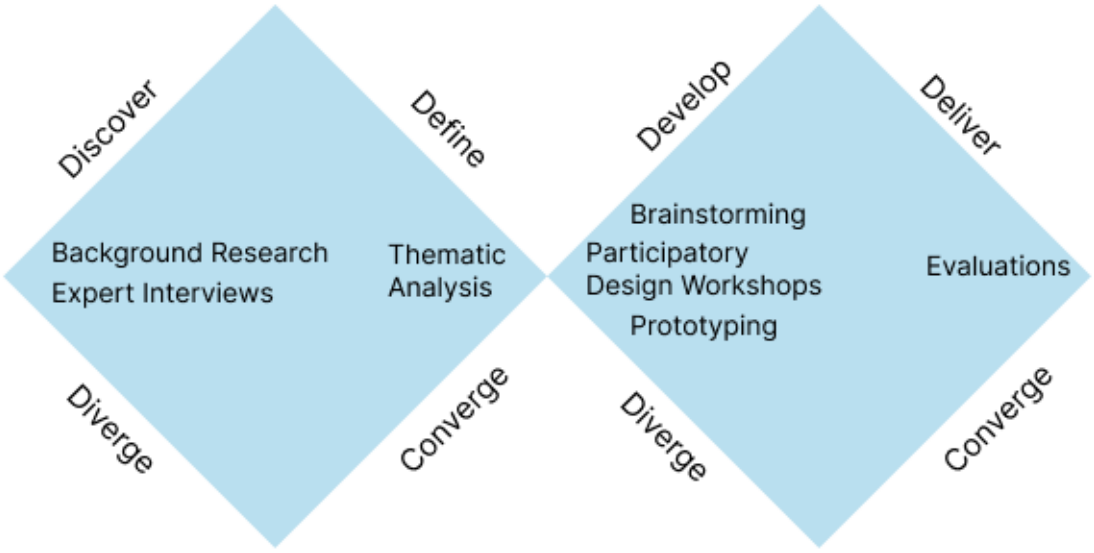


Figure 5.1: Double Diamond with the methods carried out at each phase.

5.1 Discover

The discover phase, which consists of understanding the problem space and its users, was carried out through background research and expert interviews.

5.1.1 Background Research

During the initial steps of this research, thoroughly understanding the research area was crucial. This was done by searching for important keywords on Google Scholar as well as reading the reports done by the Swedish partners in the EXPLAIN project. The previous reports made by the Swedish partners in this project were especially useful to gain deep insight into the specific industrial process we were working with and how it works in relation to XAI. From their reports, we also found additional

literature relevant to XAI. The related research highlighted in this phase mainly touches upon XAI, different approaches, common pitfalls, and the process of eliciting user requirements for XAI. Furthermore, research related to UI design, UX design, and gamification provided relevant background to the design work in later stages, as each area of research feeds into the design work. A summary of the background research can be read in Chapters 2 and 3.

5.1.1.1 Previous Research in the EXPLAIN Project

The previous research by the Swedish partners in the EXPLAIN project directly fed into our design work. We used the user requirements already defined by them as described in Section 2.3 in Chapter 2 together with the new requirements elicited from our interviews with the operators. Furthermore, their previous field studies were videotaped and served as observational input for the upcoming design work. Through these we received valuable additional insights, complementing our research.

5.1.2 Expert Interviews

Through expert interviews with the operators, we aimed to gain further insight into the process they work with and to identify further user requirements for the IET components. Due to logistical circumstances, the interviewed experts were operators in the bleaching process, but since they use a very similar system as the delignification operators that we are designing for, they still provided valuable feedback. Furthermore, the interviewed bleaching process operators had previous experience with an AI recommendation system that had an integrated feedback feature, similar to the IET components we are designing. This provided us with valuable insights into what their experience with that system was like, what worked and what did not work. In total, we managed to interview five operators. While the number of expert users is low, we believe it is sufficient to get insights into their work and their experience with AI systems since all the operators work in the same environment and do the same type of work. Furthermore, the goal of the interviews was not to gain a wide range of insights, but rather a deeper understanding of their experiences [107]. Several studies have been successfully carried out with a small number of expert users (5-10 people) [106, 108, 109] and thus we argue that the low amount of expert users does not affect the quality of the results.

5.1.2.1 Creating Interview Questions

Before the interviews, we prepared by doing in-depth research about the operators' field of work. This research helped us frame the problem and create interview questions that were relevant to their working context. The interviews were divided into four parts, focusing on different areas of importance.
The first part of the interviews was designed to get a general understanding of the operators' work environment and tasks. For example, what a normal workday looks like, when they take breaks, how the process they work with works, what computer systems they use, how it is used, and benefits or drawbacks with that system (see

Appendix A for more). These questions are important to ensure a UCD approach, as the IET components need to be well-adjusted to their work environment.
The second part was focused on motivation. Specifically, we created questions to understand whether the operators were more or less affected by *intrinsic* or *extrinsic* motivation in their work, based on the theory by Ryan & Desi [23]. These questions touched upon challenges and opportunities to improve within their work, why certain features in their computer systems were used more than others, and what aspects of their work they find meaningful. Through these questions, we hoped to gain insight into how and what motivates the operators in their work, and to understand how the IET components could fit within that context. More examples of these questions can be seen in Appendix A.
The third part of the interviews was focused on feedback. Specifically, as mentioned in Chapter 4, we grounded the question using theoretical approaches to XAI. For example, we utilized the connection between XAI and social sciences according to Miller [95], and created questions aimed at eliciting feedback requirements for the IET components via the operators' interactions with colleagues. For example, how they would give feedback to their colleagues, what type of feedback they would give, how they learn from each other, and potential challenges in giving feedback. We also asked about their impressions of giving feedback to digital systems rather than their colleagues. Further examples of these questions can be seen in Appendix A. Framing the questions in this manner allowed us to tackle the rather abstract concept of the IET components through the real-life experiences of the operators. Just like the operators, the XAI implemented in their workspace has to learn and improve through communication with other operators. Another benefit of this approach is that we also gain insight into how the operators prefer to work and give feedback, serving as an entryway for brainstorming and eventually a well-adjusted design of the IET components.
The fourth and last part of the interview focused on their AI experience, as some of the operators had used a previous XAI dashboard in their workspace which also had a feedback feature. These questions were therefore aimed at understanding their experience of that XAI dashboard implementation, what they thought of it, why or why not it was successful, and their impressions of giving feedback to that system. Through these questions, we hoped to gain a further understanding of what constitutes a good implementation of IET components. Examples of these questions can also be found in Appendix A.

5.1.2.2 Interview Procedure

The participants were given a consent form upon the start of the interview. After this, they received a paper where they filled out demographic information. We had five participants where four were between the ages of 30 and 39 years old and one was between 18 and 29 years old. All the participants were men, had varying degrees of education, worked as either process operators or process technicians and had varying years of experience in their current position. For the full demographic information, see Table 5.1. The interviews were semi-structured, meaning that we asked questions outside of the interview protocol when deemed necessary. The semi-structured nature of the interviews became especially useful once we discovered that

some operators had previous experience with AI dashboards and feedback features similar to the IET components. Asking more in-depth questions regarding this on the go was invaluable for the upcoming design stages. The interviews were recorded with offline recording equipment and were conducted with five operators inside the control room of a paper mill. The interviews lasted between 30-40 minutes and the questions for the operators can be found in Appendix A. The interviews were conducted in Swedish and only the quotes shown in Section 6.1 in Chapter 6 were translated into English. The interview data was then transcribed and prepared for a thematic analysis, as described in the following section.

Participant	Age	Gender	Education	Job	Experience
P1	18-29	Male	Technical college engineer	Technician	7-10 years
P2	30-39	Male	High school	Operator	1-3 years
P3	30-39	Male	Higher vocational education	Technician	4-6 years
P4	30-39	Male	4-year high school	Operator	11+ years
P5	30-39	Male	4-year high school	Technician	7-10 years

Table 5.1: Expert interview participants' demographic information. Technician and Operator refer to the same job, in the text, we call them operators.

5.2 Define

The define phase of the double diamond, which focuses on converging the knowledge gained in the discovery phase, was done by completing a thematic analysis of the data collected from the expert interviews. The insights from this part were then used for the next stage of the double diamond to feed into the design solutions.

5.2.1 Thematic Analysis

For the thematic analysis, we followed the six steps of thematic analysis as defined by Braun & Clarke [101]. The thematic analysis began with transcribing the audio data from the interviews, allowing us to familiarize ourselves with the data. To analyze the data and find codes and themes that are interesting we used the software MAXQDA 2020 [100]. We began by coding the first interview together to make sure that we were aligned in how we selected codes. After that, we continued to code the remaining four interviews separately. We selected the codes first based on the interview questions (deductive codes) and then added codes that had emerged from the interviews (inductive codes). Some quotes were coded multiple times as they fit into multiple different codes.
When all the interviews had been coded we printed out all the quotes with their corresponding codes. We then grouped the codes to find themes and sub-themes. Firstly, all the quotes were grouped according to their code. If a quote had several codes attached to it, it would be printed out the corresponding amount of times and grouped into each collection of codes. This initial grouping led to a large number of collections, so in the following step, we read through each collection and refined the grouping by placing quotes and codes into larger themes. This meant that

quotes and codes from different collections would end up in a larger more general theme. Most of these themes naturally emerged during this stage, but some themes were deductive in origin. This initial construction of themes can be seen in Figure 5.2. However, as specified in step four of Braun & Clarke [101], the initial themes need to be double-checked to make sure they represent the dataset. With this in mind, several themes were combined, removed, or renamed to better represent the dataset. During this stage, we noticed that some themes were more central, and other themes offered additional supportive insights to these themes. Wanting to keep both, the central themes were made into major themes, and the others emerged as sub-themes. Following this step, we had our themes, but step five as defined by Braun & Clarke [101] remained, meaning that the themes required further refinement. At this point, the quotes and codes inside each theme were summarized as bullet points, These bullet points allowed us to understand holistically what data each theme contained, which enabled us to specify the names of the themes and to better understand the entire data set. The four major themes that emerged were *Human to AI Communication*, *AI to Human Communication*, *Trust*, and *Motivation to Give Feedback*. These themes will be further described in Section 6.1. Finally, once we had our final themes, we created Fig. 6.1 representing all themes and sub-themes. In Section 6.1 in Chapter 6, we support each theme and sub-theme with direct quotes from the operators.

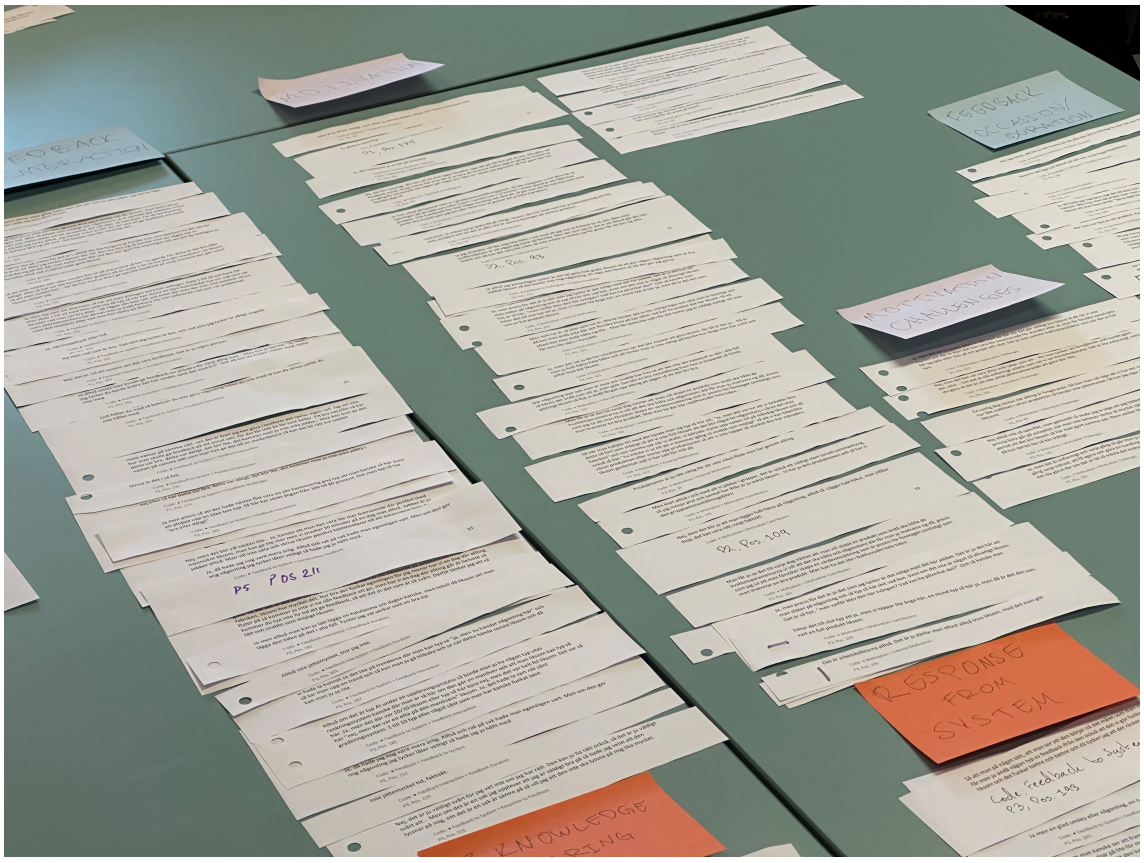


Figure 5.2: Constructing initial themes from the printed out codes.

5.3 Develop

The *develop* phase was where the design work began. We conducted multiple brainstorming sessions to use the insights gathered from the operators to come up with design solutions. Following this, we ran two participatory design workshops, the first one with UX designers from ABB, followed by a second session with operators from the paper mill. Lastly, we prototyped the solutions in preparation for evaluations.

5.3.1 Brainstorming

Brainstorming was the initial step in the design iterations, which was done by using the method Brainwriting 6-3-5 as described in the below section. We chose to work with a graphic method of brainstorming that includes both sketching and writing, as it would provide a broad set of design ideas. This initial brainstorming session was conducted with the two researchers of this thesis, mainly to explore the design space in preparation for the upcoming design workshops.

5.3.1.1 Brainwriting 6-3-5

We conducted Brainwriting 6-3-5, where we wrote or sketched three ideas for five minutes before switching the papers and building on each other's ideas. This method is quite flexible and while it is recommended to do this exercise with six people, we thought it was sufficient to do it with two people as an initial exploratory step.
The two main components of our work are about how to motivate the operators to give feedback to the AI system, and the design of the IET components. These two things were therefore the focus of our brainstorming sessions.
One of the ideas derived from this brainstorming method was to utilize several AI models producing different predictions which the operators could choose from. One idea was to visualize the different AI models as advisors on a team. We thought that utilizing a metaphor would make it easier to visualize the concept we are trying to convey while also making it more intuitive and fun for the operators to use.
Another set of ideas was about how the operators can provide feedback to the system. In a way, the selection of an AI model gives direct feedback to the data scientists about which models are useful. Another idea was that the operators should also be able to rate the predictions based on how much they trust them and how certain they are of the prediction accuracy.

5.3.2 First Participatory Design Workshop

The first participatory design workshop was conducted with four UX designers at ABB, some of whom are a part of the EXPLAIN project [5]. The session aimed to generate more design ideas for the IET components and took around two hours. We intended to use multiple methods such as Brainwriting 6-3-5 [73], rapid prototyping [72], think-aloud [84], and dot voting [110]. However, due to time constraints, we only conducted Brainwriting 6-3-5 and rapid prototyping. However, we still got the outcome we were hoping for; additional ideas and input for the design.

5.3.2.1 First Participatory Design Workshop Participants

At the start of the session, the participants filled in a consent form, detailing how their data would be gathered and used for the research. They also filled in their demographic information. There were six participants, two of whom were the authors of this thesis. There were three men and three women, where three were UX designers and one was UX team manager. The three UX designers had between one and three years of experience in their current role, and the UX team manager had between four and six years of experience. Four of the participants were part of the EXPLAIN project [5]. All the demographic information can be seen in Table 5.2.

Designer	Gender	Education	Job	Experience	Part of EXPLAIN
D1	Female	B.Sc.	Student	N/A	Yes
D2	Male	B.Sc.	Student	N/A	Yes
D3	Male	M.Sc.	UX team manager	4-6 years	Yes
D4	Male	M.Sc.	UX-designer	1-3 years	Yes
D5	Female	M.Sc.	UX-designer	1-3 years	No
D6	Female	M.Sc.	UX-designer	1-3 years	No

Table 5.2: First participatory design workshop demographics, showing the designers' demographic information. Participants D1 and D2 are the authors of this thesis.

5.3.2.2 First Participatory Design Workshop Procedure

A protocol was created for the workshop which was sent out to the participants before the workshop to read through in their own time, see Appendix B. It was also handed to them at the start of the workshop so that everyone could see the purpose of the session and be aware of what steps they were going to complete. The session started with an introduction of the topic of this thesis and a summary of what had been found thus far concerning motivation and how the operators could give feedback. The participants got to read through the workshop procedure (see Appendix B) and ask any questions they might have. Following this, we started with the brainstorming activity which was Brainwriting 6-3-5 [73]. In this exercise, each of the six participants got a sheet of paper that was folded into three sections. Everyone was asked to sketch or write down three ideas (one in each section of the paper) for five minutes. When the five minutes had passed, the participants passed their paper clockwise to the person next to them. In the next round, they were asked to add to the previous participant's ideas. This procedure was repeated five more times until everyone had added ideas onto everyone's papers and each person received their original paper back. Each person looked through everything that had been added to their original ideas before presenting each of the ideas together with the additions made from the others, as seen in Fig. 5.3.
Since the brainwriting 6-3-5 took longer than we accounted for, we decided to continue the session with rapid prototyping [72]. At this point, two of the participants had to leave and the remaining four continued. The participants split into two groups with two people in each group and spent 20 minutes creating rapid paper prototypes

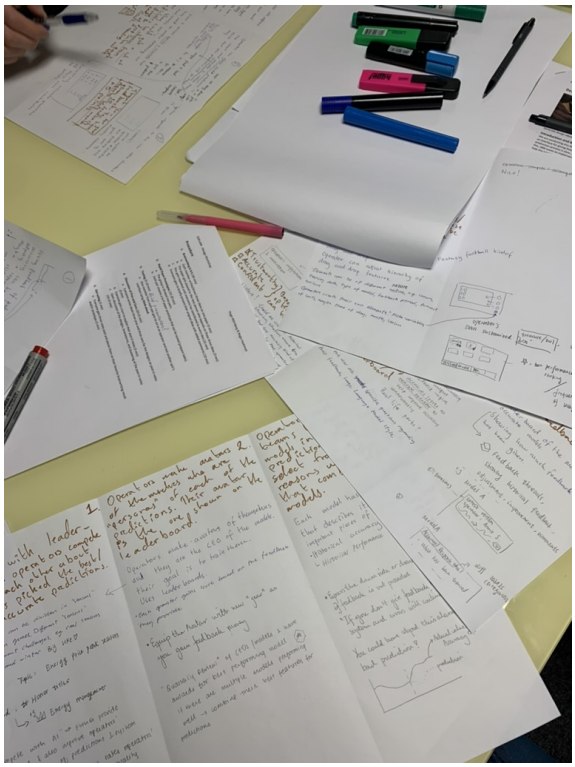


Figure 5.3: Outtake from the participatory design workshop.

of one or multiple ideas from the brainwriting 6-3-5 session. When the prototypes were completed, they were presented and discussed among the participants.

5.3.2.3 First Participatory Design Workshop Analysis and Results

As these are very initial results and do not produce any substantial insights in and of themselves, we chose to present the results of this first workshop here. Especially since the following initial prototyping step takes great inspiration from the ideas of this workshop. In total, 17 ideas emerged as a result of the brainstorming session. These were read through and grouped based on similarities, in total four groups were created which are presented below.

- **Interactive elements:** Three of the ideas were related to incorporating some form of interactive elements in the design. Summarized, they state that the design solution should be interactive, meaning that the operators can edit values, move lines in graphs, use drag and drop, utilize sliders, and other similar features to work with the predictions and their feedback.
- **Gamification:** Six of the ideas were related to incorporating some form of gamification, serving the purpose of increasing motivation. Common design options were scoring systems with leaderboards and in-game rewards such as badges and trophies. Another suggestion was a competition between the predictive capabilities of the AI versus the operators.
- **Visualization of variables:** Three of the ideas were related to some form of visualization. For example, visualizing the variables that the models consist of is important and was highlighted in many of the designs. Some suggestions

were to use radar charts and percentages to visually compare models based on their variables. One idea touched upon increasing trust in the AI by visualizing their reasoning and comparing its resulting prediction with other sources.

- **Comparing Models:** Five of the ideas related to giving the operators the ability to compare several different AI models with different predictions. This requires differences to be properly visualized between models and for the models to be distinct enough to produce different predictions.

The rapid prototyping step resulted in two designs. The first design focused on a gamified way of representing different models using metaphors. Each model is represented by a character with corresponding characteristics, such as being cost-effective or liberal. The second idea was also related to gamification. Here, the operators would produce their own predictions and compare their accuracy with the AI model. Each operator would get a score representing their overall accuracy, and they would compete against each other by giving accurate feedback to the AI.
Based on these insights from the first participatory design workshop, we began the initial prototyping stage.

5.3.3 Initial Prototypes

After the brainstorming session and the first participatory design workshop, we began the prototyping work as part of the development phase of the double diamond. Following the rapid prototyping process [72], we began with sketches, one of which can be seen in Fig. 5.4. These sketches were mainly based on the ideas from the brainstorming sessions. During this initial design work, we made sure to cross-check the pre-defined requirements for the IET components (see Chapter 2), as well as the additional insights identified from the thematic analysis (see Chapter 6).

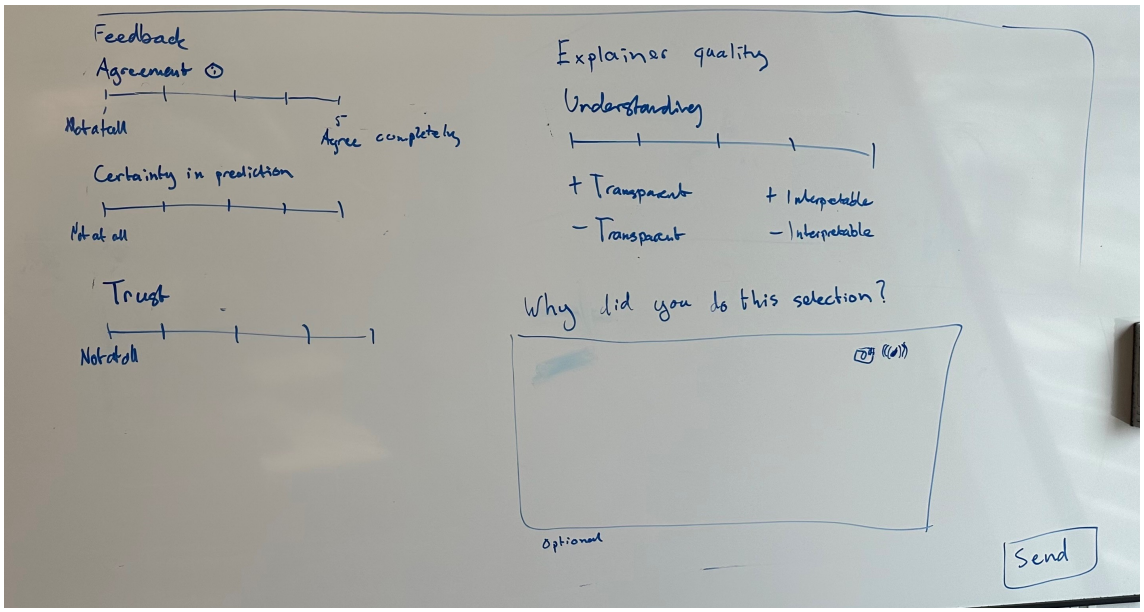


Figure 5.4: Initial prototype sketch.

The initial prototype, as seen in Fig. 5.4, is mainly focused on figuring out what type of feedback the operators value. Since these initial sketches created the founda-

tion for the upcoming participatory design workshop with the operators, we could try different approaches and validate or discard them in the participatory design workshop. For example, do they want to give feedback on the estimated accuracy of the prediction, the interpretability of the explanations, or their trust in the prediction? For this initial design, we utilized various sliders to gauge how the operators felt about the prediction and the quality of the explainer. We also imagined radio buttons for two of the characteristics of explainability quality: transparent and interpretable, as we think that the explanation can either be transparent or not, and interpretable or not, and it is not something that could be rated on a scale. We also added a box for free-text answers that could be used to supplement the feedback given through the sliders. In short, the feedback method in these initial sketches did not differ much from the previously used AI system the operators were familiar with. However, since it includes various nuances of feedback, it would provide insight into what type of feedback the operators value most in the next stage.

5.3.4 Prototypes for Second Participatory Design Workshop

Following the first participatory design workshop and the initial prototyping, we translated the paper sketches into digital low-fidelity prototypes created in Figma [74] in preparation for the second design workshop where the basics of the design were showcased but the functionality was limited. These prototypes were printed on paper and used as material for the second design workshop. These prototypes do not differ functionality-wise from the initial sketches and are simply refined for improved communication of the ideas. One idea, however, was only loosely defined in words. This idea would enable the operators to directly interact with the prediction by adding hypothetical "what if" sensor data and actions into it. Furthermore, we created two versions of gamification, where the operators can gather points based on the feedback they give. The first version was a competitive version where the operators competed against each other, and the second version was collaborative where all operators gathered points as a team and worked towards milestones and rewards. These can be seen in Fig. 5.8. Some parts of the XAI dashboard, unrelated to feedback, had been designed by ABB at this point. Since these designs were crucial parts of the XAI dashboard and gave necessary context to the IET components, we created low-fidelity versions of these designs and printed these out as well. All the designs created at this stage, including the loosely defined idea, were used as material for the second participatory design workshop.

5.3.5 Second Participatory Design Workshop

Before the second participatory design workshop, we created low-fidelity prototypes of the IET components and the XAI dashboard based on the insights from the first participatory design workshop. We provided the operators with printed-out versions of these components (Fig. 5.5, 5.6, 5.7, and 5.8), allowing them to freely move them around and discard or sketch new design suggestions as needed. The goal of the session was to understand what features for providing feedback the operators preferred. Although the participating operators were primarily bleaching

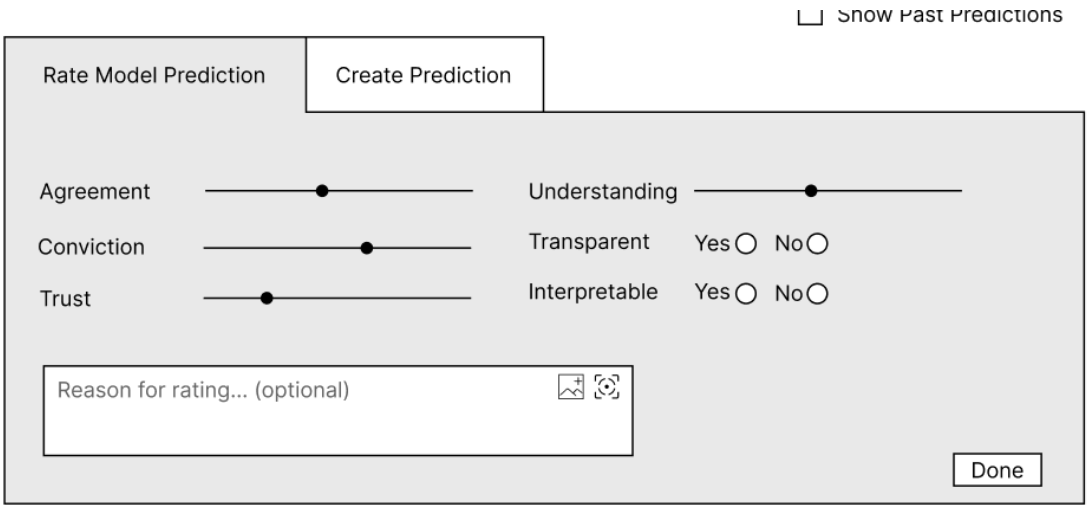


Figure 5.5: One alternative to the *Rate Model Prediction* feedback view.

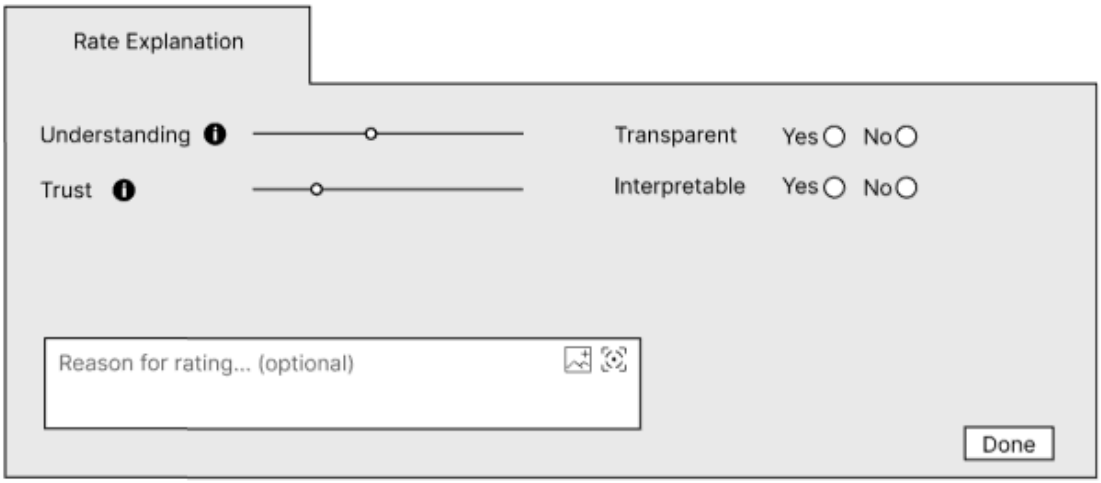


Figure 5.6: Second alternative to the *Rate Model Prediction* feedback view with a smaller set of parameters.

operators, we argue that they provide generalizable insights into the delignification process. Especially since some operators have diverse work experience from both the bleaching process as well as the delignification process, we argue that the participants possess a nuanced view of the design problem, also effectively contributing to a multi-stakeholder perspective and generalizable insights.

5.3.5.1 Second Participatory Design Workshop Participants

Six operators were part of this workshop, two female and four male. Three participants were between 18 and 29 years old, two were between 30 and 39, and one was between 40 and 49. They had varying degrees of education and worked as process technicians or operators. Two participants had worked in their current role for one to three years, one had worked for four to six, two had worked for seven to ten, and one had worked for eleven or more. For full demographic information, see Table 5.3.

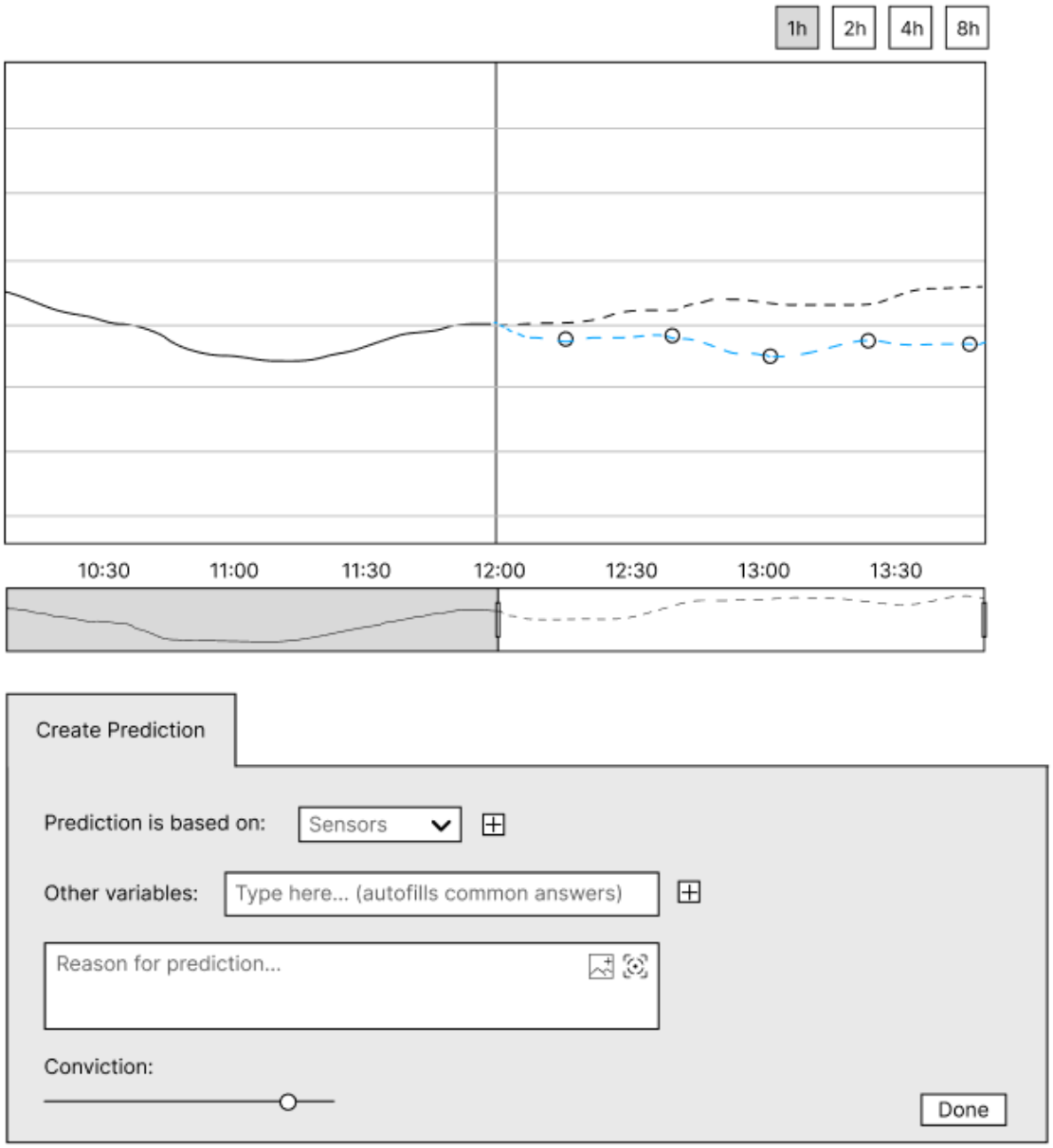


Figure 5.7: *Create Prediction* view with an interactive graph and the feedback window where the users get to select what sensors the prediction is based on.

5.3.5.2 Second Participatory Design Workshop Procedure

We conducted three sessions with the operators of the paper mill. Each session consisted of two operators and two designers (the authors of this thesis). The operators each filled out a consent form and the session was recorded capturing both video and audio. The sessions' purpose was to let the operators "build" and design their ideal version of the XAI dashboard, using existing design suggestions developed by us and ABB. The session began with a short introduction by the researchers, followed by the design exercise. The short introduction mentioned the results from the thematic analysis on motivation to give feedback, allowing the operators to get into

Leaderboard

Name	Most Used Model/h	Accurate Predictions	Amount of Feedback	Total
Name 1	+10p	+20p	+20p	50p
Name 2	+10p	+20p	+20p	50p
Name 3	+10p	+20p	+20p	50p
Name 4	+10p	+20p	+20p	50p
Name 5	+10p	+20p	+20p	50p

Scoreboard

Name	Improved Precision	Accurate Predictions	Amount of Feedback	Total
All Operators	+8%	+10p	+20p	50p

Figure 5.8: Two versions of *gamification*, the table at the top shows a competitive implementation and the bottom table shows a collaborative gamification implementation.

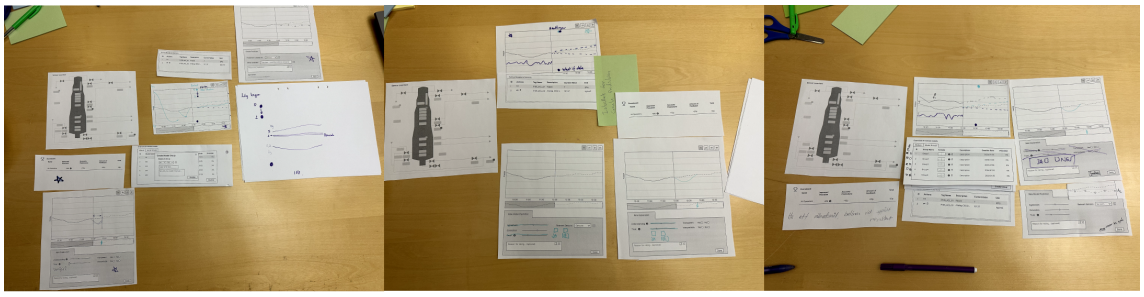
the mindset that any feedback interaction should support motivation. Regarding the design exercise, first, we presented the XAI components responsible for communicating information to the operators, i.e. how the interface explains its prediction to the users. Second, we asked them an open question: this is how the system communicates with you, how would you like to communicate back? This led to brief reflections on what type of interaction they would prefer. Gradually during this discussion, we introduced our design suggestions, asking the operators about their impressions and opinions. We also introduced ideas for including elements of gamification, with the main difference being that it was either collaborative gamification where the operators are on the same team gathering points towards milestones, or that they compete against each other for points. The operators were encouraged to add or remove features from these design suggestions to better align with their preferences. Once all ideas had been presented and discussed within the group, the operators were asked to discard any ideas they thought were unnecessary. At this stage, only the ideas the operators considered useful remained, and most of these had been slightly modified by them, as seen in Fig. 5.9.

Participant	Age	Gender	Education	Job	Experience
O1	30-39	Female	Higher vocational education	Technician	1-3 years
O2	18-29	Female	High school	Technician	1-3 years
O3	18-29	Male	Higher vocational education	Technician	7-10 years
O4	40-49	Male	High school	Operator	11+ years
O5	30-39	Male	Higher vocational education	Technician	4-6 years
O6	18-29	Male	Technical college engineer	Technician	7-10 years

Table 5.3: Operators' demographic information from the second participatory design workshop. Technician and Operator refer to the same job, in the text, we call them operators.

5.3.5.3 Second Participatory Design Workshop Analysis

The final selection of interface components was cross-checked between all operators and analyzed using affinity diagramming (see description of the method in Section 4.4.3 in Chapter 4). The Post-it notes were mainly based on observations and represented their choices in the workshop (their choices can be seen in Fig. 5.9). The occurrence of each interface component was also noted and any additions to the design added by the operators also resulted in a note. Furthermore, any interesting insights based on discussions or observations within the workshop were also noted down on post-it notes. This resulted in a large selection of notes, which were grouped based on affinity. These were then combined to create a smaller set of notes representing the sentiment of the grouping. This resulted in three major themes. The results of this affinity diagramming can be seen in Section 6.2 in Chapter 6.



G1 G2 G3

Figure 5.9: The resulting choices of interface components from the second participatory design workshop. Shown in order of Group 1, Group 2, and Group 3.

5.3.6 Refined Prototypes

In this section, the refined prototypes will be described. Both of these versions were designed by the authors of this thesis, with some components being provided by ABB. The refined prototypes were based on the results of the thematic analysis of the initial interviews, and the affinity diagramming from the second participatory design workshop, which can be seen in Sections 6.1 and Section 6.2 in Chapter 6. Since A/B testing was chosen as the method for evaluating the interface, as will

be discussed in Section 5.4.1, two versions of the dashboard were created as part of the refined prototyping step. In the following sections, all parts of the interface will be explained, both the ones relating to the IET components as well as the XAI dashboard in general.

5.3.6.1 XAI Dashboard

The entire dashboard consists of four main components: sensor overview, overview of trained models, kappa forecast, and active models and sensors. These components had already been created by ABB and our focus was on incorporating the IET components into these. All IET components are created by the authors of this thesis. In this section, each component created by ABB will be explained.
The sensor overview provides the operators with an interface similar to what they encounter in their other computer systems, aligning with their existing mental models of the process [80, 81, 82, 83]. Once a model has been chosen, the sensor overview displays the feature importance for the most prominent sensors [36]. Furthermore, it is also interactive and the users can drag and drop the sensors into other parts of the dashboard. The overview of the trained models component provides the user with a list of all available models to choose from. Each model varies in feature importance (shown in the sensor overview) and produces slightly different predictions. The kappa forecast component displays the current prediction and provides the user with the ability to navigate back in time and view previous occasions. Finally, the active models and sensors component allows the operators to display the graphs of several different AI models, as well as any sensors that they drag into this space from the sensor overview.

5.3.6.2 Version A

In this section, Version A will be described in detail.
To see a kappa forecast, the operators need to select a model from the *Overview of Trained Models* component ((2) in Fig. 5.10). An automatic model selector with the same purpose has already been explored [118], however, there is research showing that allowing users to explore contrasting features in predictions plays a role in user trust [119]. To this end, the model selection is a part of the IET components as a way of providing feedback. Since the most accurate and useful AI models will be selected more often, this becomes a form of natural selection of AI models. Once a model is chosen, the *Kappa Forecast* component ((3) in Fig. 5.10) shows a graph of the predicted kappa value where the solid line is the actual kappa value going back in time, and the dashed line is the predicted future value. Below the kappa forecast graph is a set of *actions* and *effectors* ((6) and (7) in 5.10). Both *actions* and *effectors* can be dragged into the prediction graph to change the prediction graph in real-time. Doing so enables the operators to adjust the prediction to better align it with their assumptions of what will happen. In particular, *effectors* are variables that affect the kappa, e.g. the moisture of the incoming wood chips, the season, broken sensors, sawdust, and others. For example, if they know that they will receive wood chips with more moisture soon, they can then place it into the graph and update the prediction accordingly. The operators can also drag hypothetical

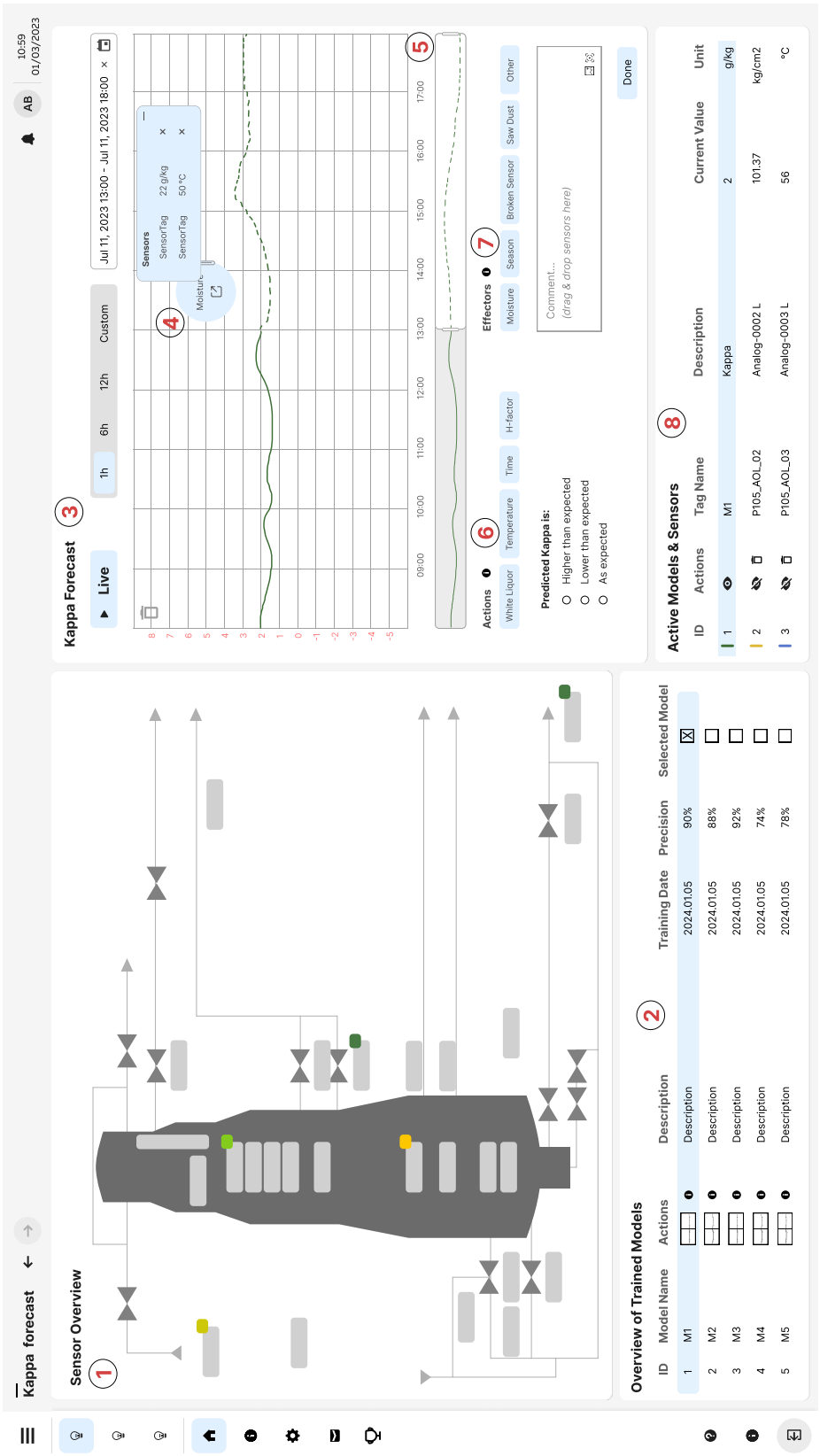


Figure 5.10: Version A of the interface. (1) is the sensor overview, (2) overview of trained models, (3) kappa forecast, (4) an effector (moisture) with the sensor window open, (5) the graph navigation tool, (6) actions to drag into the graph, (7) effectors to drag into the graph, and (8) the active models and sensors in the graph.

actions into the graph - actions they can take to ensure the process remains stable. For example, if they know that adjusting the H-factor will help make the process more stable again after they have received wood chips with more moisture, they can drag the H-factor into the graph and change the predicted value.

In Version A the *actions* and *effectors* are represented by bubbles that "push away" the graph. Where the bubble is placed affects the prediction, i.e. if it is placed close to the graph it will push it away more than if it was placed further away from the graph. In short, the bubbles do not directly affect or interact with the AI model, they simply act as obstacles for the prediction graph to avoid. This allows the operators to adjust the prediction according to their assumptions of what will happen. By clicking on the expand symbol in the bubble a window opens. The operators can drag and drop sensors from the sensor overview into the window. By doing this, they tell the AI that certain sensors are affected by the *action* or *effector* ((4) in Fig. 5.10).

The *action* and *effector* bubbles can be expanded by grabbing and dragging the grab handle icon. Making the bubbles wider indicates that they affect the prediction for longer. The standard size of the bubbles represents one hour and the operators can drag them out if they, for example, think that the effects of more moisture will be ongoing for two hours. The same principle goes for the actions - the standard size represents one hour and they can be dragged out to represent a longer period. When the time of the prediction has passed, the operators will see if the *actions* and *effectors* they dragged into the graph corresponded with what happened. If the changes in the prediction due to these additions were accurate with what happened, the bubbles will turn green. If it is not accurate, the bubbles will turn gray.

Dragging *actions* and *effectors* into the graph effectively labels the data, which is a current issue since most training data is unlabeled. Labeling the data is therefore a form of feedback. Additionally, adjusting the prediction to align it with the operators' assumptions also gives insight into how much these labels should affect the prediction. If the operators' adjustments are accurate, this is valuable feedback. For future reference, the feature of dragging *actions* and *effectors* into the prediction will be referred to as the Interactive Prediction Graph (IPG).

Underneath the *actions* and *effectors* is a feedback space. There the operators can give both short and longer feedback. The inclusion of these features can be motivated by the Fogg Behavior Model [102, 103, 104], as actions that are easier to do require less motivation to be initiated. The shortest and quickest form of feedback that they can give is the radio buttons, where they select whether they think the kappa value is higher than expected, lower than expected, or as expected. If they want to, they can also elaborate on their feedback by writing in the text box to the right, where they can also add pictures or drag-and-drop sensors that impact their feedback. The operators click done when they are satisfied and receive a pop-up window that thanks them for providing feedback. The window shows a summary of what their feedback contains, as well as an indication of how many points they collected. They can click cancel which will take them back to the previous stage where they can edit their feedback. If they click confirm the feedback is sent to the data scientists.

5.3.6.3 Version B

The main differences in Version B are how the *actions* and *effectors* interact with the AI model, as well as how the operators can provide quick and brief feedback.
In this version, similar to version A, the operators need to first select a model from the *Overview of Trained Models* component ((2) in Fig. 5.11). The *Kappa Forecast* component ((3) in Fig. 5.11) shows a graph of the predicted kappa value, the solid line is the actual kappa value going back in time, and the dashed line is the predicted future value. In this version the *actions* and *effectors* are placed above the kappa forecast graph ((4) in Fig. 5.11). The principle is the same as in version A where they can be dragged into the graph to change the prediction, but with the main difference that the added *actions* and *effectors* directly interact with the AI model and provide an updated, more informed prediction, instead of only adjusting the prediction by pushing the prediction graph away.
Also, instead of bubbles the *actions* and *effectors* are shown as boxes with a line that indicates when the *action* or *effector* will happen. When dragging an *effector* into the graph the operators specify the duration of the effect, and drag and drop sensors from the sensor overview into the box to show which sensors will be affected by the *effector*. Crucially, once the operators drag sensors into the box, they can change the sensor value to anything they desire. This enables the operators to type in hypothetical values that the sensors could have in the future ((5) and (6) in Fig. 5.11). Doing so will update the prediction with the new information, and change it accordingly. This is the main functional difference between the two versions. Version B enables the operators to directly interact with the model and to see how it reacts to any scenario imaginable. When the scenario has played out, the operators can see if the *actions* and *effectors* they dragged into the graph changed the prediction in a favorable way or not. The boxes are green if the predictions were more accurate compared to the original prediction, and gray if it was not as accurate. A text will also appear on the boxes, further specifying if each *action* or *effector* was more or less accurate than the original prediction.
Similar to Version A, dragging *actions* and *effectors* into the graph effectively labels the data, which is a current issue since most training data is unlabeled. Labeling the data is therefore a form of feedback. Direct interaction with the AI model is beneficial in two additional ways. Firstly, it allows for the exploration of potential future scenarios, making the AI model a tool for the operators to use in their day-to-day work to make more educated and informed decisions. Secondly, it functions as an explainer (see Section 2.1.1.1 in Chapter 2 for more explainers), allowing the operators to understand how much the model takes the *actions* and *effectors* into account. It allows the operators to effectively converse with the model and see how it responds to hypothetical scenarios, seeing how it aligns with their assumptions.
Underneath the kappa forecast graph is the graph navigation tool and feedback space. Here, the operators can select a period in the graph navigation tool and add comments. When clicking on the add comment symbol, a window opens where they can give feedback on whether the predicted kappa is higher than expected, lower than expected, or as expected. Again, the inclusion of these features can be motivated by the Fogg Behavior Model [102, 103, 104], as actions that are easier to do require less motivation to be initiated. They can also add a free-text comment

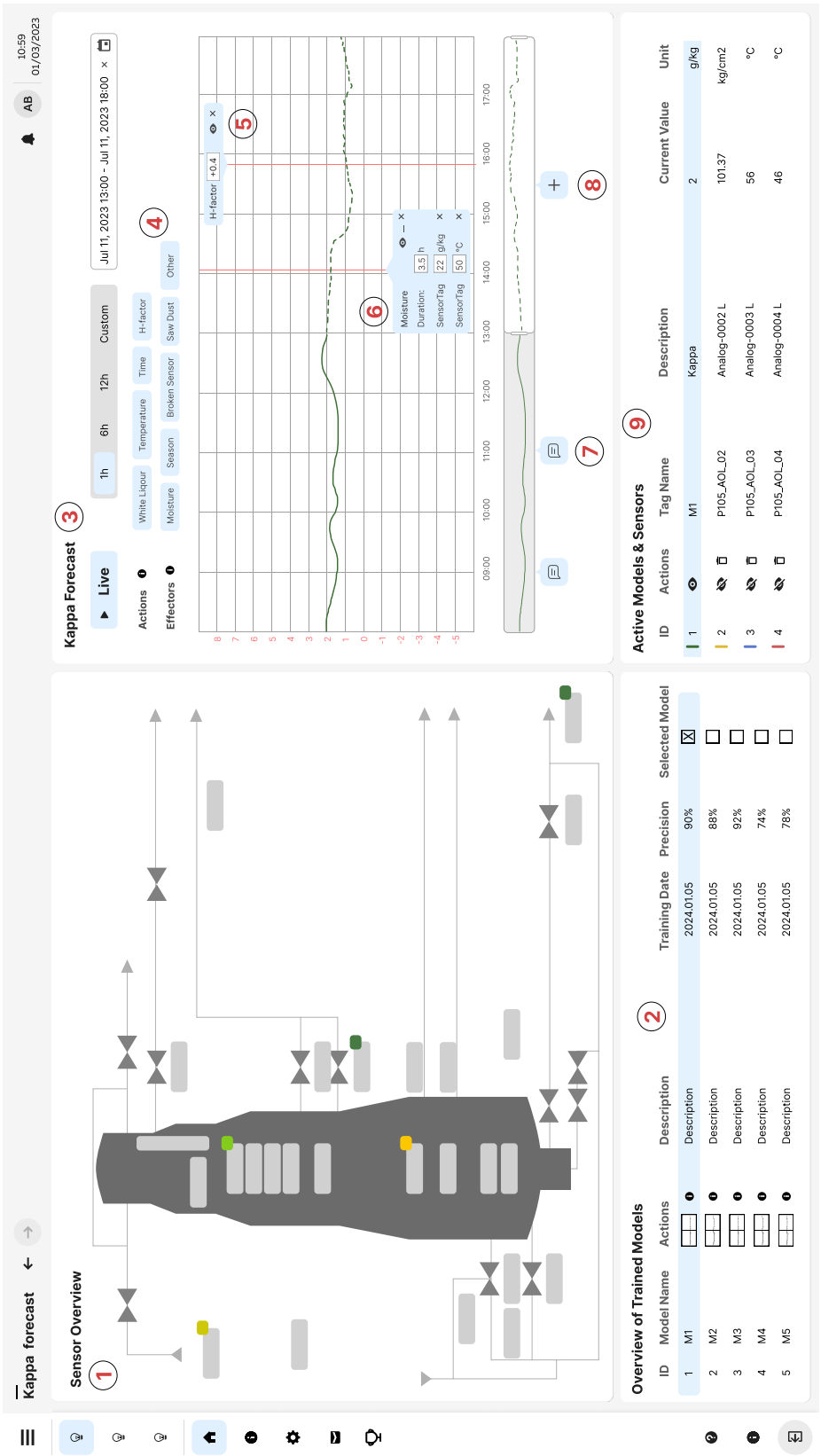


Figure 5.11: Version B of the interface. (1) is the sensor overview, (2) overview of trained models, (3) kappa forecast, (4) actions and effectors to drag into the graph, (5) an action (H-factor) in the graph, (6) an effector (moisture) in the graph with its affected sensors, (7) a previous comment shown on the graph navigation tool, (8) add new comment, and (9) the active models and sensors in the graph.

and include pictures if they want to ((8) in Fig. 5.11). Underneath the text box, there is a space for sensors to be dragged and dropped from the sensor overview, to indicate that the feedback affects those specific sensors. Once a comment has been placed, they are marked with the comment symbol underneath the graph navigation tool ((7) in Fig. 5.11). The operators can at any time open each comment to read, edit, or delete if they want to. This feature enables old comments to be reviewed or looked at for reference, which is a difference compared to Version A. Otherwise the functionality is the same.

5.3.6.4 Scoreboard

A scoreboard was also created to be used in the evaluations. The inclusion of the scoreboard is based on research on gamification where Darejeh et al. [53] demonstrated that 81% of gamified software has positive results on user engagement and motivation. Furthermore, according to Xu [56], rewards are the main motivator for gamified systems. The scoreboard is accessed through the trophy symbol in the left sidebar. The purpose of the scoreboard was to understand what gamification-related information and features the operators value. Gamification is something that was discussed with the operators in the second participatory workshop (see Section 5.3.5) and they were generally positive to the idea as long as it meant that it was not a competition between individual operators or shifts, but that all operators competed towards common milestones. The operators also expressed that they appreciate statistics and graphs, and we therefore designed a page that focuses on statistics to show how the models have improved over time and how the operators' feedback has helped the models to improve.

The scoreboard is divided into quarters of the year and the operators can choose which quarter to look at from the drop-down menu. The first part of the scoreboard page shows the milestones the operators are working towards. Each milestone is worth 50 points and there are three milestones each quarter. The operators get a reward when they have reached a milestone, for example, a "fika", which is illustrated by the coffee cups. An alternative to this was also created where the milestones change to medals, showing if they have reached bronze, silver, or gold.

Underneath the milestones, three components are showing different statistics. The first one, the amount of feedback, shows how many times the operators have provided feedback, how many points they got per feedback occasion, and how accurate the feedback they provided has been compared to the actual kappa values.

Next to the amount of feedback is a bar chart showcasing statistics for each of the models. The operators can toggle between two views, one to see the runtime of each model and another to see how much feedback each model has received.

Lastly, the scoreboard page has a graph showing how the model precision has changed over time. The graph shows three months at a time, which corresponds to the quarter selected at the top of the page. The graph shows all the models and how their precision has changed. When hovering over one or multiple lines, an overlay will show what percentage precision those models have at a specific point in time.

5.4 Deliver

In the last phase of the double diamond, *deliver*, we conducted evaluations on the interactive prototypes. The insights from the evaluation were used for a final iteration of the prototype and were also used in combination with the other results from this thesis to create design recommendations for future work with IET components.

5.4.1 Evaluations

To test the refined prototypes, three evaluations with operators from the paper mill were carried out. The evaluation procedure consisted of three main parts, going through scenarios using the prototypes with the think-aloud method [84], A/B testing [88], and finally a brief summative interview targeting the research questions. These three parts were analyzed independently, but some impressions gathered in the think-aloud session were included in the analysis of the summative interview. The results can be read in Section 6.3 in Chapter 6. The goal of the think-aloud session was to identify usability issues and gain detailed qualitative insights during usage. As mentioned, some of these qualitative insights were also included as material for the analysis of the summative interviews. The goal of the A/B test was to expose the users to alternative designs to compare multiple feedback methods in terms of motivation and suitability. The summative interviews were conducted to gather broader and more in-depth data about the features of the IET components, the gamification and scoreboard aspects, and how they relate to the research questions in this thesis. While the think-aloud session gives detailed insights into the usage of the IET components, the summative interview provides more general and holistic insights. Together, they offer valuable insights, which is why the think-aloud data unrelated to usability issues was analyzed together with the summative interview data. Furthermore, the questions in the summative interview were based on the research on XAI evaluation methods [29], as described in Section 3.2.3 in Chapter 3. The questions were mainly aimed at getting insight into user satisfaction, understanding, curiosity, motivation, and trust. Coupled with the qualitative insights from the think-aloud session, this was achieved.

5.4.1.1 Participants

The evaluations were done with three operators who each tested the interface for an average time of 45 minutes. Following a brief introduction, the operators received a digital consent form. All three operators were male, two were between 40-49 years old and one was between 30-39, two had completed high school, one had completed university, two had eleven or more years of experience, and one had between seven and ten. All the demographic information can be seen in Table 5.4. It is interesting to note that the expert users in the evaluation on average had more work experience in their current role compared to the expert users of the previous interviews and workshops. Furthermore, none of the operators had been part of the previous interviews or the workshops. Since they were entirely unfamiliar with the work in general and the interface, they offered fresh insights into the functionalities and design of the prototypes. Although it can be discussed that the number of

Participant	Age	Gender	Education	Job	Experience
E1	40-49	Male	High School	Operator	11+ years
E2	40-49	Male	High School	Operator	11+ years
E3	30-39	Male	University	Operator	7-10 years

Table 5.4: Evaluation participants' demographic information.

expert users is low for generating significant results, it has been argued that three participants can be sufficient for identifying most of the usability issues [113].

5.4.1.2 Evaluation Procedure

The evaluations were done remotely using Microsoft Teams. The full protocol for the evaluation sessions can be seen in Appendix C. Each session started with an introduction to the purpose of the session and how it would be carried out. Each participant also filled in a consent form, detailing that the data would be collected using recordings of their voice and screen, as well as how it would be used in the context of the research. They then received a link to the Figma prototype and were first asked to read through the interactive walk-through of the interface where each component was explained.
After familiarizing themselves with the interface, the participants were instructed that they would go through several hypothetical scenarios in the interface and to "think-aloud" [84] during their interaction with it, meaning that they vocalized and described everything that they were doing and why they were doing it. Both version A and version B of the interface had identical scenarios, although with different instructions since the interactions varied slightly between the two versions (see Appendix C for scenarios). Due to these differences, version B had 3 scenarios and version A had 2. For each version, we started with the quickest and easiest scenarios before building into the longer more complex scenario.
Following the completion of the scenarios for versions A and B using the think-aloud method [84], the participants were asked which version they preferred and why. To reduce bias, participants one and three started with version B and participant two started with version A.
Finally, a brief summative interview was carried out specifically targeting their impressions of the interface, whether it is useful, and how it relates to their motivation to use the interface. The full set of questions can be seen in Appendix C.

5.4.1.3 Think-Aloud Analysis

The overarching goal for the think-aloud method was to extract usability issues and detailed insights for the IET components. As previously mentioned (see Section 4.5.1.2 in Chapter 4), the think-aloud method allows for multiple qualitative analysis methods [115]. In this case, we opted for a deductive coding framework, labeling interactions as successful or unsuccessful, with the goal of revealing and understanding any usability issues in the interface. The entire session was transcribed, including annotations for where the users clicked and interacted with any part of the interface. The data was then imported into Figma as Post-it notes. Once imported,

all data was sorted according to its corresponding scenario and interface version. Successful interactions were assigned a green-colored Post-it note, and unsuccessful interactions were assigned a red-colored Post-it note. The entire sorting was cross-checked between the two researchers of this thesis. The majority of the interactions were successful, however, since the goal of this analysis is to extract usability issues, the unsuccessful interactions were highlighted and presented as results. These can be seen in Section 6.3.1 in Chapter 6. Importantly, not all data gathered from the think-aloud session was analyzed in this way. The insights unrelated to usability issues, concerning the impressions of the interface were analyzed together with the data from the summative interviews.

5.4.1.4 A/B Testing Analysis

After the participants completed the scenarios, they were asked which version, A or B, they preferred and why. These results can be seen in Section 6.3.2 in Chapter 6.

5.4.1.5 Summative Interview Analysis

The goal of the summative interviews was to extract both broader and more in-depth insights into the operators' impressions of the functionalities of the IET components and more specifically the IPG feature. To gain this sort of data, we extracted the relevant impressions from the think-aloud session and analyzed it together with the rest of the interview. Another important feature related to motivation is the gamification aspects of the design. These were mostly present in the scoreboard part of the interface, and some questions were aimed at understanding the operators' impressions of these features concerning rewards, points, and AI improvement statistics. The main aim of this thesis is to understand what increases the operators' motivation to give feedback to an AI through IET components. Thus, the questions in the evaluations were aimed at understanding how the functionalities of the IET components relate to increased motivation to use the IET components. To target this, the questions about increased motivation were linked to their intent of using the IET components. The full set of questions can be seen in Appendix C.
The collected data was analyzed using a thematic analysis as described by Braun & Clarke [101]. First, the data was transcribed. The transcribed data was then coded and the codes coupled with the quotes were grouped to understand the overall patterns of the data. Due to the small amount of data, this coding and grouping could be done manually in Figma, and no other software was used. The codes were mainly deductive, a result of the questions we asked, however, some inductive codes naturally emerged as well. Similar to the previous thematic analysis (see Section 5.2.1), we followed the six-step process as defined by Braun & Clarke [101].
Due to the low number of participants and resulting interview dataset, the process of analyzing the summative interview data was simple compared to the previous thematic analysis in Section 5.2.1. Almost all of the themes were deductive since they were a result of the interview questions.
The initial themes were double-checked to ensure that they represented the dataset and were combined, removed, or renamed where it was fitting. Following this, relevant quotes in each theme were checked to make sure they represented each

theme accurately. After the themes were created, a figure was made to represent the themes, sub-themes, and their connections to each other. This figure, together with direct quotes from the evaluations, can be found in Section 6.3.3 in Chapter 6.

6
Results

This chapter showcases the results from the *define*, *develop* and *deliver* phases of the double diamond [66]. From the *define* phase, we show the results of the thematic analysis. From the *develop* phase, we present the results of the affinity diagramming, and from the *deliver* phase we show the results of the conducted evaluations.

6.1 Thematic Analysis of the Expert Interviews

The result of the thematic analysis, as described in Section 5.2.1, is presented in Fig. 6.1. In the figure, the circles represent the themes and the rectangles represent the sub-themes. The lines represent the connections between the themes and sub-themes. The themes will be presented with a description of how they connect to each other in the following sections, followed by the sub-themes and relevant quotes from the transcripts. As all themes and most sub-themes are interconnected and tie into each other, a holistic interpretation of the thematic analysis is necessary to understand the results.

6.1.1 Themes

Four major themes emerged as a result of the thematic analysis. As each theme connects to the others as well as several other sub-themes, they and their relation to each other will be generally described in this section. The next section on sub-themes will delve deeper into the specific take-aways from the interviews.

6.1.1.1 Human to AI Communication

This theme highlights some of the most important considerations concerning how the operator interacts with the AI. This theme contains several interconnected sub-themes which are individually described in the next section on sub-themes, which can be seen in Fig. 6.1. The connected sub-themes mainly touch upon how the operators wish to communicate with the AI and some difficulties in communicating about the process they work with. Some of the connected sub-themes also overlap with how they wish for the AI to communicate with them, which generally demonstrates brief, direct, and continuous communication. In general, many of the operators talked about how human-to-human communication differs from human-to-AI communication. Several operators expressed that human-to-human communication requires social sensitivity, while human-to-AI communication should be more straightforward. This is further discussed in the sub-theme *Continuous, Brief, and*

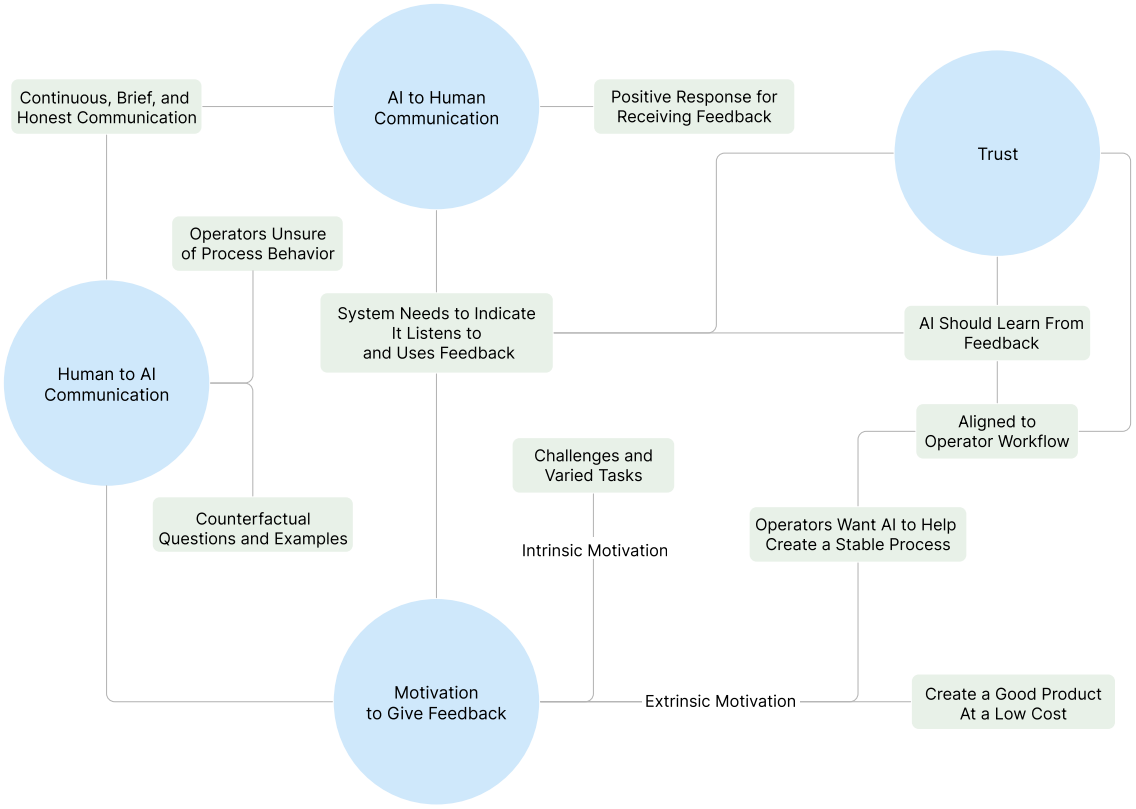


Figure 6.1: Figure representing the results of the thematic analysis of the expert interviews.

Honest Communication. However, the operators reflect that it would be good if the interactions with the system were similar to those with other humans, but without the social sensitivity.

> *[...] if you can give it almost the same way as you can give it to a colleague, then it can be quite fun [...] (P2)*

Thus, utilizing existing manners and characteristics of communication, while still maintaining a direct and concise nature, could be beneficial for engaging the operators' motivation to give feedback.

6.1.1.2 AI to Human Communication

Similarly to how there are considerations for how the operators interact with the AI system, there are also considerations regarding how the AI system interacts with the operators which is shown in this theme and its interconnected sub-themes. This theme and the *Human to AI Communication* theme which is described above shares the sub-theme *Continuous, Brief, and Honest Communication* as that type of communication is something that is required from both parties to have a successful interaction, which can be seen in Fig. 6.1. The AI needs to communicate with the operators in a way that is understandable and easy for them, as they have stated that they do not want to spend all their working time communicating with the AI. Many of the operators also wished that the AI would respond in a kind and positive manner, something that mirrors most of their interactions with their

colleagues. When asked how they would like the AI system to respond to feedback, one operator said:

> *The system should maybe ask a counter question and say ... or give a suggestion of something to do instead, for example. (P1)*

The other sub-themes in this theme are related to *motivation*. This comes from the operators saying that the system needs to show them how it listens and uses the feedback they provide, because if they do not see that their feedback is valued, they will stop using the system and stop giving it feedback. These are factors that feed into their instrinsic motivation for continuously using the system. Not only should the AI communicate that it listens to and uses feedback, but also demonstrate that it improves over time. Demonstrating this trait is greatly related to increased trust.

6.1.1.3 Trust

As shown in Fig. 6.1, *trust* is related to *AI to human communication*. For the interactions between the AI and operators to be successful, the operators need to feel that they can trust the AI. It can be summarized as: if the operators trust the system, they are more likely to be motivated to provide feedback. When asked about what is important when implementing an AI system similar to the one they previously used, one operator said:

> *You want to trust it. Because if we are to follow it all the way (the AI's recommendations), then we should do that and you should be able to trust it. (P4)*

As discussed in Sections 3.2.1 and 3.2.2 in Chapter 3, trust is paramount for XAI systems to succeed. It is important that the system is aligned with the operators' workflow and demonstrates improvement since it will increase trust in the system. For the operators to trust the system, the communication between the AI and the operators needs to meet their requirements. The system should learn from the feedback the operators provide and show that it has learned via effective communication, to keep them motivated in continuing to use the system.

6.1.1.4 Motivation to Give Feedback

Several factors influence motivation to give feedback. This theme emerged as central in the thematic analysis and is deeply interconnected, showing that motivation to give feedback is dependent on a multitude of variables (as seen in Fig. 6.1). Firstly, as described in the theme *AI to Human Communication* and its interconnected sub-themes, the operators stated that it is important for them to know that their feedback is being used and that the model indicates that it both listens to and actively uses their feedback. Failing to indicate *how* their feedback is received and used by the model negatively affects their motivation to give feedback. Secondly, the motivation to give feedback is affected by the types of interactions necessary to do so. As described in the sub-theme *Continuous, Brief, and Honest Communication*, interactions should follow these principles to ensure ease of use. More time-consuming feedback results in repeated monotonous and drawn-out interactions which has a negative impact on motivation to give feedback. Thirdly, motivation to provide feedback is affected by the operator's trust in the model. As

described in the theme *Trust* and its interconnected sub-themes, the operators need to know *how* the model works and improves. As is shown in Figure 6.1, trust is highly connected to the theme *AI to Human Communication*, which in turn connects to motivation to give feedback. If they trust that the model improves because of their feedback, they will be more motivated to give feedback to it. When asked about what encourages them to give feedback to an AI, one operator said:

> *[...] a lot of the people I talk to here think that AI is the next big thing. So I think there is a lot of people who are motivated to help to see how good it can be. Because it could offload a lot in stressful situations or in potentially dangerous situations where you work with high pressure and temperatures. (P3)*

Lastly, as can be seen in the connecting line *Extrinsic Motivation* between *Motivation to Give Feedback* and *Create a Good Product At a Low Cost*, the operators are determined by their goal of creating a good product for the company they work for. Having an AI model that assists in making that process more stable and easier to manage emerged as a general motivation for the usage of AI. Thus, gradually improving the AI is in the best interest of the operators, which demonstrates a pre-existing baseline of motivation to give feedback. However, this baseline of motivation is only as useful as the model itself, since the operators lose motivation to provide feedback if the model fails to acknowledge or use it.

6.1.2 Sub-Themes

In this section, each sub-theme will be presented alongside relevant quotes from the transcript. As some sub-themes are interconnected to other themes, their relation to those will be highlighted as well.

6.1.2.1 Operators Want AI to Help Create a Stable Process

During the interviews, the operators mentioned that they lost motivation to provide feedback to the previous AI system partly because it was not integrated into the other software used to control the bleaching process. Since the system was isolated, it did not have any ability to act on its own, and it had to be constantly manually updated on the ongoing process. This meant the system gave vastly different outputs each time it was updated, leading to many ups and downs in the process. In turn, this led to the operators gradually losing trust and motivation to use the system.

> *[...] you got new answers every time. But in our everyday life, we don't make so many changes in such a short time. And they probably wanted to try to find a way to get everything in the process stable, but it turned into a roller coaster instead. (P1)*

Given that the new AI system developed by ABB will function in the same way as the previous system used, it is important to consider how the IET components and XAI dashboard could simulate either hands-on operations or improvements to its decision-making, without tampering with the process or the model itself.

6.1.2.2 Continuous, Brief, and Honest Communication

This sub-theme emerged as a result of how the operators communicate, learn from, and give feedback to each other as well as how they wish to communicate with an AI system. They emphasized the importance of an ongoing open dialogue with positive undertones with their colleagues, as a way to keep morale high and to benefit from their colleagues' expertise. P5 stated that to ensure that the process runs smoothly, cooperation and open dialogue are required between colleagues.

> *I expect that we should try to agree on a way forward. If my idea is better, or this person's idea is better. It is oftentimes very difficult to know and so we need to communicate [...] (P5)*

To some degree, this also applies to the AI system, as it needs to be up-to-date with what is going on in the process. However, continuous communication with the AI must be carefully considered, as it is important that communication with the AI system does not require unnecessary attention. It should strike a fine balance of being informative while not being disruptive. The operators state that they prefer brief segments of communication to an AI and that when it came to giving feedback, the majority of them agreed that if giving feedback was too time-consuming, it would be problematic. Concerning brief communication to the AI system, P1 stated:

> *To a system it should be brief. After all, it is not a person. (P1)*

To ensure that the operators learn from each other's expertise, they must communicate in a brief and honest way. P3 stated that:

> *I expect that people are honest, so that if I ask about something I want them to tell me what they think. Not that they are too agreeable or try to be nice, but that they are as constructive as possible. So that we help each other improve. (P3)*

Similar to how colleagues communicate with each other, P5 argues that interaction with an AI should follow these principles to an even larger extent.

> *I would be more honest (to an AI). More direct. But if it gives me something I think seems reasonable, I would agree. (P5)*

As highlighted in the theme *Human to AI Communication*, these two quotes demonstrate a difference in interacting with AI systems compared to interactions with their colleagues. This sub-theme shows that communication in all forms to and from the AI should be direct and honest to ensure that it maintains substance.

6.1.2.3 Operators Unsure of Process Behavior

Since the bleaching process is rather complex, involving several variables unknown to both the operators and the computer systems, it is important to remember that the feedback given by the operators cannot always be accurate. Some situations in the process are simply different and cannot be understood with the available data.

> *It is very complicated to determine how effective the bleaching will be. Sometimes we don't even know. We measure some things, and other things that matter, such as metals or impurities in the mass, are impossible for us or the computer to know. (P5)*

6.1.2.4 Counterfactual Questions and Examples

Counterfactual questions and examples emerged as a sub-theme based on how the operators currently work and solve problems. Mainly, while encountering unexpected problems, the operators tend to use trends in their data to backtrack its origin or to "trend back" to similar data points where the problem was solved. Trends allow the operators to both diagnose and solve problems. Examples of other similar situations to the current state could be explored both as communication to and from the AI. Furthermore, P1 explicitly stated that counterfactual questions could be useful to enhance the usefulness of the AI:

> *The system could give me a counterfactual question and give a suggestion as to what would happen if we do this instead. (P1)*

6.1.2.5 System Needs to Indicate It Listens to and Uses Feedback

For the operators to be more motivated to use the AI system and give it feedback, it must indicate in some way that it both listens to the feedback they provide, as well as how it is used. Many of them said that they stopped using the previous AI system because they felt that it was not listening to the feedback they gave it and thus not improving its recommendations:

> *If I had worked with an AI and you didn't notice a difference within a year, you might ask yourself "Does what we're doing even help?". But if you work with an AI where you every three months get a small update you might notice that "yes it's improved since January" so you would be more motivated to continue giving it feedback and continuing to improve it. Because if it takes too long you might ask yourself "What are we even doing? Why are they even receiving feedback?". (P3)*

Regarding listening to feedback, an important factor is to demonstrate that the AI improves. The operators said that one of the drawbacks of the previous AI system was that the system itself did not receive the feedback, but the developers did, who then implemented it into the model.

> *If it had learned by itself, then it surely could have developed faster, but now it was only those who programmed that got the feedback and had to change it afterward. (P5)*

This is similar to how ABB's system will function as well. Thus, to avoid demotivating the operators, transparency about how the model will be incrementally updated by data scientists is important. Not only should the system indicate that it listens to feedback, but it should also be able to show *how* the feedback is being used. An interesting suggestion to address the issue of both indicating listening and usage of feedback is to directly simulate improvements to the AI based on the feedback.

6.1.2.6 Challenges and Varied Tasks Are Motivating

Many of the operators mentioned that the most fun and rewarding parts of their work are when there are challenging situations and they need to work to solve problems. These situations often allow them to learn more and they therefore find them to be more fun than when the process is stable and all they need to do is monitor it.

When asked about what makes them lose motivation, one participant said:

> *Monotone tasks. We don't have a lot of those, but it can happen. That's the good thing about this job that it isn't so monotone and there are lots of different things every day, so that's part of the fun, that you can choose a lot of different tasks. (P1)*

This affects *intrinsic* motivation as they enjoy work more when there are challenges and opportunities for them to develop in their work. This sub-theme should be taken into consideration for the way in which the operators provide feedback.

6.1.2.7 Create a Good Product At a Low Cost

This sub-theme is a large part of the operators' *extrinsic* motivation. The operators' *extrinsic* motivation is driven by the company's goals, which in turn affects their motivation to use the current system they use, as well as their future uses of an AI with the IET component. This sub-theme demonstrates that their goals in the process of creating a high-quality product are paramount to their motivation to use AI in their work. When asked about their goals with the bleaching process, many of the operators mentioned that their goal is to create a high-quality product at a low cost. This *extrinsic* motivation is what drives them to perform their job well.

> *Every day you make sure to create a product that has the quality parameters we want it to have. Somewhere there you have to motivate yourself, and you try to create a positive revenue for the company while delivering a good product. You can keep that in mind a bit at all times. (P2)*

6.1.2.8 Positive Response for Receiving Feedback

Many operators mentioned that when they give feedback to each other they hope to get positive responses from their colleagues:

> *When I give feedback I think I would like a positive reaction about giving feedback. (P2)*

Given the operators' attitudes towards giving feedback and the theories presented in Section 3.2.1 regarding how human-to-human interactions can be applied to human-to-AI interactions, we can conclude that if the operators prefer to receive a positive response from colleagues, they also want the AI system to respond positively. The same participant mentioned the following about how they would like the computer system to respond:

> *Polite. A polite, positive response. (P2)*

6.1.2.9 AI Should Learn from Feedback

The operators talked about how for them to use the feedback functionality more, they needed to see that the system was learning from the feedback they had provided. Knowing that the AI improves in response to feedback is a large contributor to *extrinsic* motivation. Some of the operators mentioned that because the previous AI system they had used was not learning from their feedback, they want future systems to do so:

> *So that you in some way can see that it is reaching the goal you had put up, maybe. Then you still get some type of feedback from it as well that what we are doing works, it learns more and it works better and then I think that's enough. (P3)*

6.1.2.10 Aligned to Operator Workflow

Some of the operators mentioned that there were issues with how the previous AI system they had used was not aligned with their workflow. The previous system gave recommendations every 15 minutes which was too seldom according to some operators, and too often according to other operators. One operator who thought that it happened too often stated:

> *Because here it was small or big changes during a very very short time, every 15 minutes. And a bleaching sequence you get step-by-step responses after 1 to 2 hours. And we made changes during that period so we were not entirely sure what we had indirectly touched. So I would rather have had a function where you suppress it for hours before it did a change. (P1)*

This shows that the previous AI system was not aligned with the way the operators worked and thus it affected how they worked with it. This sub-theme is connected to the *Trust* theme, through *Intrinsic Motivation* and *Extrinsic Motivation*. When the system is aligned with how the operators work, they trust the system more, while it also internally motivates them to use it. The operators are also externally motivated to use the system more when it aligns with their work as it means that it assists them in running the entire process more smoothly.

6.2 Second Participatory Design Workshop - Affinity Diagram

This affinity diagram is the result of the second participatory design workshop where the operators had a chance to build their own XAI dashboard with feedback functionality. As mentioned in Section 5.3.5 in Chapter 5, the operators could choose from a set of interface components created by us and by ABB. They were tasked with choosing or discarding these components, as well as discussing important functional considerations when providing feedback to an AI system. Thus, the resulting affinity diagram is backed up with quotes from the operators.

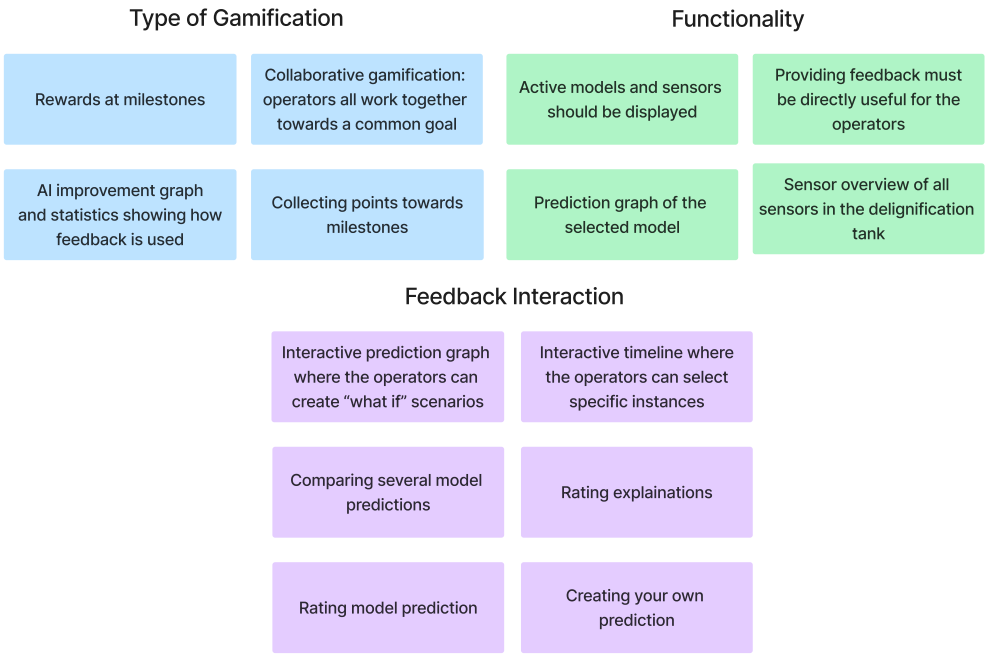


Figure 6.2: Analysis of the results from the second participatory design workshop.

As seen in Fig. 6.2, the operators had some preferences regarding feedback interaction. They wish the timeline itself to be interactive so that specific instances can be selected and interacted with to give feedback. The popularity of the different components varied. For example, rating the model prediction and creating your own prediction to compare with the models were added, but its usage over time was debated. The operators argued that since there is no instant benefit to providing this type of feedback, interest in providing it would wane over time.

> *Perhaps it could be fun in the beginning, but I think that you would get tired of it after a while because you do not directly benefit from it. (O2)*
> *This type of feedback is only useful for the AI. If it is to be used in the long run it needs to be useful for us running the process as well. (O3)*

To address this issue, feedback has to be directly useful at the moment. To this end, the operators preferred the idea of making the prediction graph itself interactive. Doing so enables the operators to place "what if" sensor data into the prediction, which recalculates the prediction including the updated sensor data. Further expanding this idea could also enable the operators to place hypothetical actions into the prediction. This doubles as feedback since the prediction would be updated with new information and an explanation of what will change and why it will do so. It also becomes a tool for exploring future scenarios and allows the operators to prepare and plan ahead. Outside factors unknown to the computer, such as weather and moisture impact the kappa value, and if the operators could import the relevant sensor data into the prediction to account for this information, the prediction becomes more accurate.

> *You get direct benefit from this quite quickly. And afterwards, you can*

> *see if you were correct or not, compared to if you send your feedback to somewhere where you do not know where it ends up. (O2)*
>
> *With clear graphs I think it is really good. (O4)*

Another suggestion that was popular among most operators was the ability to choose between several models with varying feature importance and predictions. Over time, natural selection allows the best-performing models to stand out by comparing runtime. This also becomes feedback allowing the models to improve.

Furthermore, as seen in Fig. 6.2, all operators agreed that a collaborative gamification solution is preferred. The reason is that any other form of gamification, whether competing against each other individually or between shifts, would become problematic in the long run. When asked about gamification focused on the operators competing against each other, either between shifts or between individuals, the operators were generally against the idea.

> *Sure, I am a competitive person, but I do not think this is suitable here. I really do not think so. We are all working towards the same goal. (O3)*
>
> *It could create a divide between shifts. (O6)*

Furthermore, giving feedback to the system is mostly done when the predictions are problematic, which oftentimes corresponds with problems in the process itself. Since the process could be stable for long periods, it would be an unfair competition. Collaborative gamification was much preferred since all operators work towards the same goal of keeping the process stable. A few suggestions for this that arose during the session was a scoring system where the operators work towards milestones and rewards. Many operators also stressed the importance of displaying the impact their feedback has on the models by showcasing statistics of improved precision over time. The operators agreed that such statistics, rewards, and milestones would improve motivation to provide feedback.

Concerning functionality, as seen in Fig. 6.2, the operators agree that the sensor overview is essential, allowing them to understand what the models account for in their predictions. Naturally, the prediction graph and the ability to plot specific sensor data into that graph should also be present. Another functional insight that was discussed previously is the importance of making the act of giving feedback directly valuable and useful for the operators. The operators argue that interest in providing feedback will diminish over time if it is only useful for the model. These insights are essential to ensure that motivation to provide feedback remains high over time.

6.3 Evaluations

In this section, the results from the evaluations of the two versions, version A (Fig. 5.10) and version B (Fig. 5.11) will be presented. The evaluation was analyzed in three parts: the think-aloud results, A/B testing questions, and summative interview questions. Each part will be described in its dedicated section below.

6.3.1 Think-aloud Results

The think-aloud results were analyzed by reading the transcripts from the recordings and identifying the usability issues, as described in Section 5.4.1 in Chapter 5. Overall, the interfaces were highly usable and only a few instances stood out where usability issues surfaced. Only the usability issues are shown in Table 6.1. In the table, the function is first presented together with A or B in parenthesis to indicate whether the function was presented in version A, version B, or both. This is followed by a description of the interactions where there were usability issues in a numbered list and a statement corresponding to the usability issue.

Function	Interaction	Statement
IPG (A)	1. Struggles to expand duration 2. Adding content 3. Struggles to interpret post-event prediction accuracy	1. *"It's not logical, I would need to know that this could be done." (E1)* 2. *"If I add a bunch of things at the same time, that would be a bit cluttered." (E3)* 3. *"I can't clearly see what the prediction showed previously, only what the actual value was." (E1)*
IPG (B)	1. Clicks on wrong component to add sensor 2. Struggles to interpret post-event prediction accuracy	1. *"The broken sensor then?" (E3)* 2. *"It would be good to see what the model thought originally. [...] I want to be able to compare the previous prediction with the actual outcome." (E1)*

Comment (A)	1. Struggles to add sensors to comment 2. Not seeing previous comments 3. Confirmation box impression	1. *"I thought it would be in effectors, but I probably missed how I was able to do it..."* *(E2)* 2. *"I want to be able to see old comments [...] It's good to learn from each other and to think in the same way."* *(E3)* 3. *"I think it's a bit unnecessary."* *(E3)*
Comment (B)	1. No exact time span while viewing previous comment 2. Not seeing the original prediction within old comments	1. *"There is no time though, only the highlight in the graph. It would be good to see it more accurately."* *(E3)* 2. *"This trend looks good so I don't know why they put higher than expected here."* *(E3)*
Adjust Prediction (B)	1. Difficulties in creating their own prediction	1. *"It's difficult for us in practice to know what it's going to look like in two hours, we might fool the model instead of helping it. "* *(E1)*
Two Graphs (AB)	1. Struggles to add new model	1. *"Ah, I was on the wrong side. I see it now."* *(E2)*

Table 6.1: Usability issues that surfaced from the think-aloud method.

When participants used the IPG functionality in version A, some struggled with how to expand the duration of the added *effector* moisture. Some had issues with adding content to the graph, and some had problems interpreting the post-event prediction accuracy, i.e. the operators did not receive a satisfying response regarding the effect and outcome of their added *actions* and *effectors*. This was solely communicated via color in both versions A and B, where green represented that their added *actions* and *effectors* had a positive effect on the prediction and grey represented the opposite. Since the issue of communicating this was reoccurring for both versions, perhaps

an alternative method of displaying these effects would be better. As suggested by the operators when confronted by this (as seen in Table 6.1), showing the original prediction could be one alternative.
In version B, there were fewer issues with the IPG. One participant selected the broken sensor *effector* instead of dragging in a sensor from the sensor overview. Similar to version A, some had problems interpreting the post-event prediction accuracy.
The comment functionality in version A had some usability issues. Some participants were unsure which sensors they were meant to add to the comment and thought it was the *effectors* instead of the sensors from the sensor overview. One mentioned that not seeing the previous comments was not good, since they could learn from other people's previous comments. When asked about the confirmation box that appeared after completing a comment, some said that they felt it was unnecessary to have it.
In version B there were fewer usability issues with the comment functionality. Some said that the highlighted time span in the graph was not enough when viewing previous comments, and they wanted the exact time span to be visible in the comment box. Some also wanted the original prediction to be visible when they were viewing previous comments.
When it came to the adjust prediction functionality, which was only available in version B, the main issue was that the operators found it difficult to create their own predictions. This was mainly due to them not being certain about what the process might look like in the future and they were worried that their uncertainty might fool the model instead of helping it improve.
Showing two graphs at the same time, i.e. selecting two models to be displayed at the same time, had one usability issue where the operator looked for the models in the wrong place.

6.3.2 A/B Testing Results

When the participants had completed the scenarios, they were asked which version they preferred the most and why. Participants E1 and E2 preferred version B:

> *I prefer version B, I like that you can see the previous comments. (E1)*
> *I prefer B, it is more clear with the comments, that you can see previous ones, that is the main thing I'm thinking about. (E2)*

Participant E3 preferred version A:

> *I think I would prefer A, it is easy to write the comments here, but I would like to place them somewhere, so I can see what I am commenting on. (E3)*

6.3.3 Summative Interview Results

The short interview was analyzed through thematic analysis, as described in Section 4.3.1 in Chapter 4. A figure representing the results of the analysis can be seen in Fig. 6.3. The circles represent the overall themes found from the summative interviews, and the squares represent the sub-themes. In general, these results relate to the IPG functionality described in version B, which means that the IPG functionality

directly interacts with the AI model and updates the prediction with the added information. In the following sections, each theme will be described together with its sub-themes and relevant quotes.

6.3.3.1 IPG Functionality

One of the main findings from the interviews was about the IPG component and its functionality. This theme has three related sub-themes: *Flexibility*, *Efficiency*, and *Pro-activity*. These sub-themes describe different aspects of the IPG functionality and will be explained in the following sections. Overall, the participants in the evaluations were positive towards having a tool such as the IPG. When asked about their impressions of the IPG and its functionalities, one participant said:

> *That could be interesting, today we guess a lot, and if you have this then you could test things, and see where the kappa value ends up, it could be interesting. (E2)*

According to the operators, the IPG functionality provides insight into both the delignification process and the AI model itself, the related sub-themes delve deeper into this and the specific impressions of the IPG functionality. Furthermore, these three sub-themes support the notion that the IPG functionality offers insight, which is why the theme is connected to it in Fig. 6.3.

6.3.3.1.1 Flexibility This sub-theme refers to the large variety of opportunities that are offered by the IPG functionality. By changing values in specific sensors, or adding additional context to the prediction, the IPG functionality becomes a versatile tool for different purposes. The following are quotes related to flexibility lifted from the trancription.

> *[...]there are seasons and that is very good, many things that could be used" (P2)*
>
> *I think its good, today we could have an instrument that diverges from its actual value, then you could easily drag this is in and explain the diversion, otherwise the prediction would be wrong without this. Would be good to tell the system that it should always take this diversion into account, which is a good feature to get a better prediction. (E3)*

6.3.3.1.2 Efficiency Efficiency refers to the benefits of the graphical nature and interaction modalities offered by the IPG functionality. When using the system, two of the operators expressed that giving feedback through the IPG functionality was an efficient way of providing the model with a lot of information in a short amount of time. E3 stated:

> *It is graphically very easy, and you could do fast adjustments and fast comments that save a lot of time. [...] I like the graphical picture of the digester and that you could easily drag them into other parts. I suppose it will be more detailed in the future. I think that was good, the easy way to drag in all necessary information (E3)*

E2 related the feedback in the IPG functionality to the previous AI system they had used in their workplace:

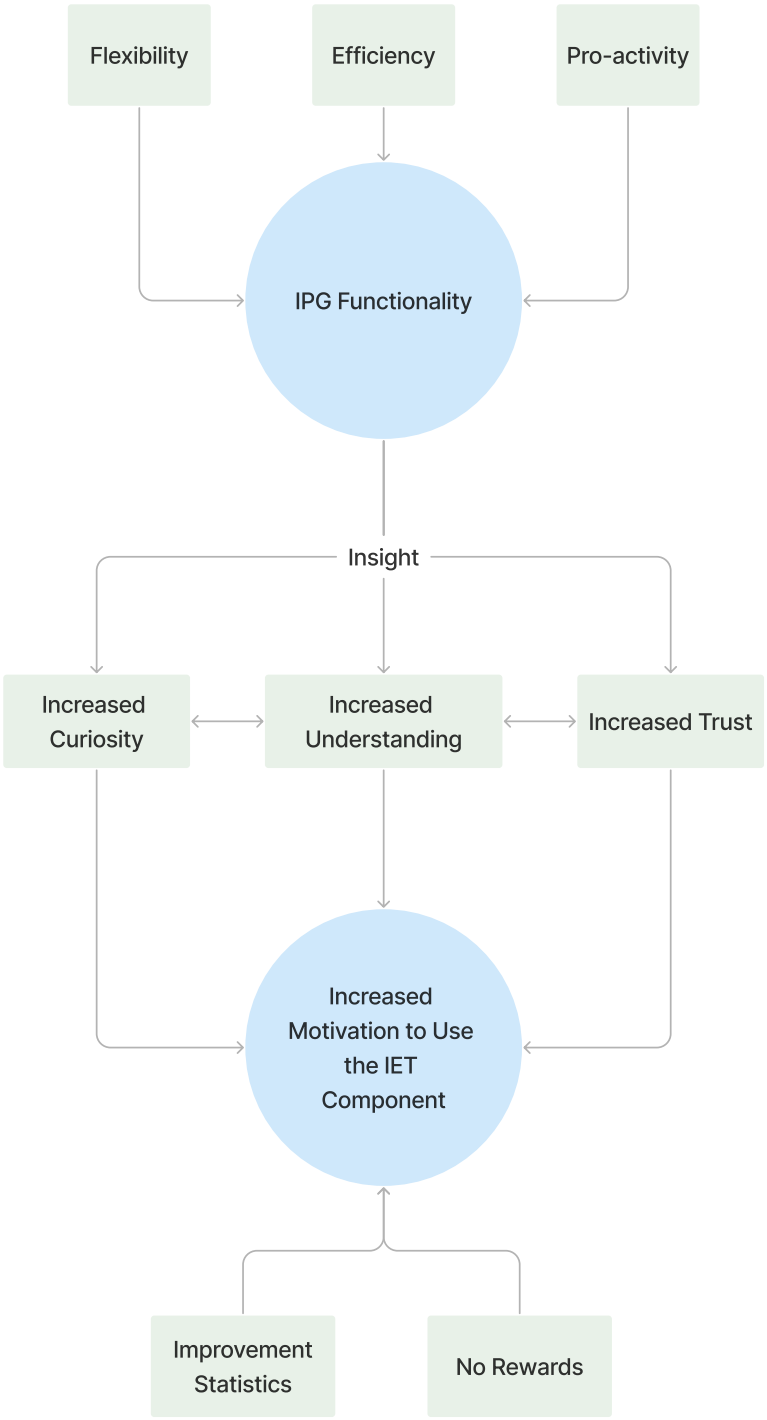


Figure 6.3: Figure representing the results of the thematic analysis of the evaluation interviews.

The previous AI, we had to write everything which took a lot of time. (E2)

6.3.3.1.3 Pro-activity Pro-activity refers to the exploratory capabilities of the IPG functionality. The first thematic analysis (Section 6.1 in Chapter 6), revealed that the operators continually communicate with each other to agree on the best action going forward since it is very difficult to know how the process will respond. The sub-theme pro-activity is related to this, and E3 stated:

Oftentimes we predict things on our own and with this, we can be proactive and see what would happen if we adjust the H-factor or add a few grams of alkali. Would be really nice to see suggestions of what the result of these actions could be. Could be important. (E3)

6.3.3.2 Increased Motivation to Use the IET Component

The next major theme that emerged in the thematic analysis was *Increased Motivation to Use the IET Component*. This theme is connected to five sub-themes *Increased Curiosity*, *Increased Understanding*, *Increased Trust*, *No Rewards*, and *Improvement Statistics*. The theme *IPG Functionality* is connected to the three sub-themes *Increased Curiosity*, *Increased Understanding*, and *Increased Trust* since the IPG functionality offers insight, which in turn leads to these connected sub-themes. The main aim of this thesis is to understand what increases the operators' motivation to give feedback to an AI through the IET components. Thus, the evaluations were aimed at understanding how the functionalities of the IET components and the inclusion of gamification elements relate to increased motivation to use the IET components. During the evaluation, the questions about increased motivation are linked to their intent of using the IET components. The operators were asked in the evaluation if the interface they tested and if any of the added gamification elements would make them use the IET components more. All three operators stated that the IPG functionality would contribute to them using the system more. When asked to motivate their response, they refer to the benefits of the IPG functionality outlined in the sub-themes *Flexibility*, *Efficiency*, and *Proactivity*. When asked about the gamification elements, the operators were unified in disliking the rewards and points, but thought that the improvement statistics were useful. This will be explained in further detail in the following sections where all five sub-themes connected to *Increased Motivation to Use the IET Component* are described.

6.3.3.2.1 Increased Curiosity The first sub-theme connected to *Increased Motivation to Use the IET Component* is *Increased Curiosity*. This sub-theme emerged as a result of the evaluations as the participants were asked if they thought a function such as the IPG would contribute to increased curiosity. E1 stated:

Yes it probably would (E1)

E2 had a similar sentiment, but underlined the importance of the predictions being accurate:

If it works then yes I would say so (E2)

6.3.3.2.2 Increased Understanding This sub-theme concerns the operators' understanding of how the AI produces its predictions. The results of the thematic analysis give insight into the operators' understanding of the AI's predictions increases with the use of the IPG. Since they can use the IPG functionality to try out hypothetical scenarios, the operators reason that it is a good way to learn how the different models will behave. When the participants were asked in the evaluations if the IPG functionality led to an increased understanding of the models used for the predictions, they stated:

> *Yes, of course, I get some more insight and I can affect it myself, which I learn a lot from. (E3)*
> *Yes I think a little bit, you learn from mistakes, and I think you would learn from it. (E2)*

6.3.3.2.3 Increased Trust Trust has been a central part when designing the IET component. During the evaluations, the operators were asked whether they think the IPG functionality would increase their trust in the system. The results indicate that the accuracy of the predictions is the most crucial factor for trusting the system. E2 stated:

> *Yes as long as it works, the previous system did not work, otherwise yes. If it doesn't work then you get a negative impression. (E2)*

E1 stated:

> *Yes I think so, I think that in the beginning you would play around with it to get a feel for the model, if the predictions are accurate, you learn to trust the model. (E1)*

6.3.3.2.4 No Rewards This sub-theme relates to the addition of gamification elements within the design. The results show that some, but not all, elements of gamification was speculated by the operators to lead to an increased motivation to use the IET component. This sub-theme emerged as a result of the operators strong negative opinions towards collecting points. When asked about the rewards, to which they all agreed that there was no point in collecting points, E2 stated:

> *I do not know if rewards in this way are good, you would feel like you would give feedback just because you should. But I think some things you should do because of interest. So for gathering points, I feel like it is perceived as a bit weird. I am skeptical to this. (E2)*

E1 was very skeptical towards any points system and stated:

> *Totally meaningless [...] It would not contribute to using the system, I think this could make you look at the model less serious. (E1)*

6.3.3.2.5 Improvement Statistics This sub-theme also relates to the addition of gamification elements within the design. The operators did have specific requirements for how they wanted the gamification elements to look and function. When asked about their impressions of the statistics that showed how the model had improved based on their feedback, the operators were more positive towards showing statistics only, as E2 stated:

> *I think this is interesting, to see how much it has improved. I think the improvement is interesting, to see how good it works, if my thoughts are accurate compared to the statistics. Seeing the change at different levels of detail is good. (E2)*

Concerning statistics, E3 stated:

> *These things are good to know, and to understand how the models has performed over time. I think the statistics are enough. (E3)*

6.4 Final Design

Following the evaluations, which were described in Section 5.4.1 in Chapter 5, we iterated on the prototypes one more time to create a final design. Since Version B was the most well-received version of the interface and the IPG functionality described in Version B was so positively received, the final design builds upon Version B, see Fig. 6.4. The IET components have all been baked into the XAI dashboard. This section will explain the core functionalities of the feedback features within the IET components and how these have been updated according to the results of the evaluations.

6.4.1 Natural Selection of AI Models

The first IET functionality is baked into the overview of trained models component. Here, the operators can choose from a selection of AI models (Fig. 6.5). The choice of models is one type of feedback, as the usage of inaccurate models gradually decreases over time and are eventually replaced by other models. This is a kind of natural selection for successful models. Since this functionality worked well in both version A and B, it has not changed in the final design.

6.4.2 Interactive Prediction Graph (IPG)

The second IET functionality is baked into the kappa forecast component (Fig. 6.6. Within the kappa forecast component that displays the current prediction, *actions* and *effectors* have been added. *Actions* are actions the operators take in their day-to-day work to ensure a stable kappa value, and *effectors* are both variables concerning the sensors as well as variables outside of the sensors that affect the kappa value. These are represented as individual interactive buttons which can be dragged into the prediction graph. These affect the prediction in two ways, either they are coupled with sensors and adjust the prediction accordingly, or they are additional AI models more sensitive to variables such as moisture. During discussions with the developers at ABB, the exact method of implementation would affect how this works. Regarding this functionality in general, most of these variables are unknown to the computer system, but known to the operators (such as moisture of incoming wood chips), and this feature enables the operators to simultaneously inform, improve, and label the prediction. Updating the prediction with additional information serves three purposes regarding feedback and motivation:

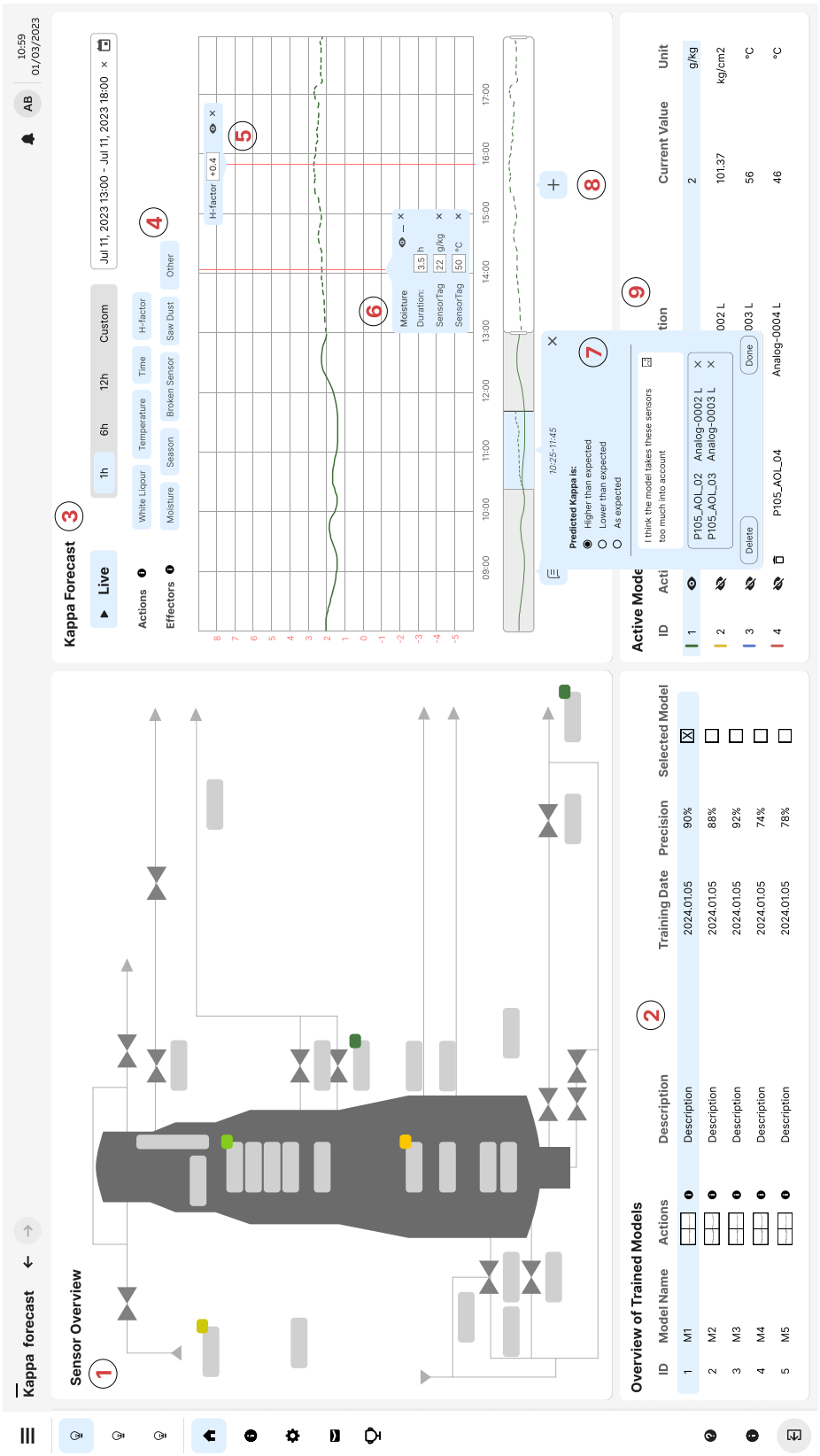


Figure 6.4: Final version of the interface. (1) the sensor overview, (2) overview of trained models, (3) kappa forecast, (4) actions and effectors to drag into the graph, (5) an action (H-factor) in the graph, (6) an effector (moisture) in the graph with its affected sensors, (7) a previous comment opened on the graph navigation tool, (8) add new comment, and (9) active models and sensors in the graph.

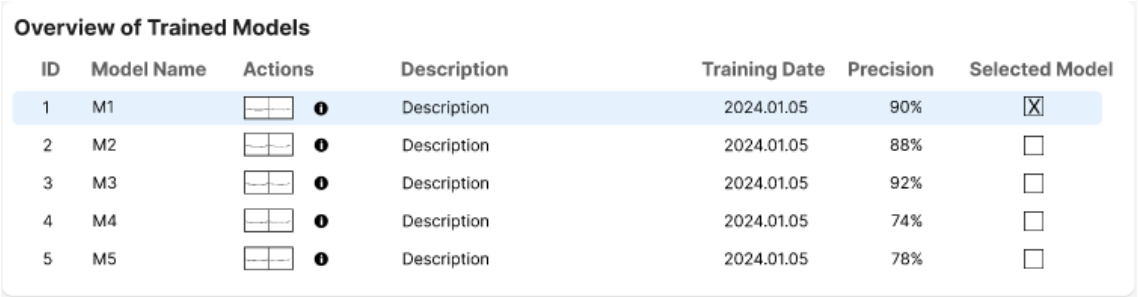


Figure 6.5: Overview of Trained Models component.

- Firstly, it serves as a form of feedback, explaining both what affects the kappa value, and why it affects it. This also effectively labels the data, which is a current issue since most of the training data is unlabeled.
- Secondly, it provides the operators with a proactive tool where they can freely ask "what if" questions, effectively giving them insight into potential future scenarios and actions. In short, they can use this to inform their future decisions and in preparation for any potential actions. Based on the results from the second participatory design workshop, seen in Section 6.2 in Chapter 6, this functionality should also positively affect their motivation to provide feedback to the system, as it directly assists them in their work.
- Thirdly, it functions as an explainer (for more examples of explainers, see Section 2.1.1.1 in Chapter 2), allowing the operators to understand how much the model takes the *actions* and *effectors* into account. It allows the operators to effectively "poke" the model and see how it responds to hypothetical scenarios, to see how it aligns with their own assumptions. Coupled with additional feedback methods in the IET components (which will be explained below), the operators can specify whether the prediction is adjusted too much, too little, or just the right amount.

The final design of the IPG is based on Version B of the interface, as two out of the three operators preferred this version. Since the results from the think-aloud protocol (see Section 6.3.1 in Chapter 6) showed that the operators struggled to interpret the post-event prediction accuracy, the design has been updated with a toggle switch enabling the operators to show the previous prediction at the time of entering the content into the IPG (Fig. 6.7. This directly addresses the wishes of the operators as they stated that it would be good to see what the model thought originally. This, coupled with an updated message in the IPG functionality showing an exact percentage of how more or less accurate their additions made the prediction (Fig. 6.7 and 6.8), should provide enough information to enable informed IPG interactions in future scenarios.

6.4.3 Dynamic Comments

The third IET functionality is baked into the graph navigation tool (Fig. 6.9). Based on the results from the interviews and the second participatory design workshop (see Sections 6.1 and 6.2 in Chapter 6), the operators need the ability to provide quick, brief, and easy feedback for specific instances of a prediction. As mentioned, the in-

Figure 6.6: Interactive Prediction Graph (IPG) component with an added action (H-factor) and effector (moisture) to the prediction.

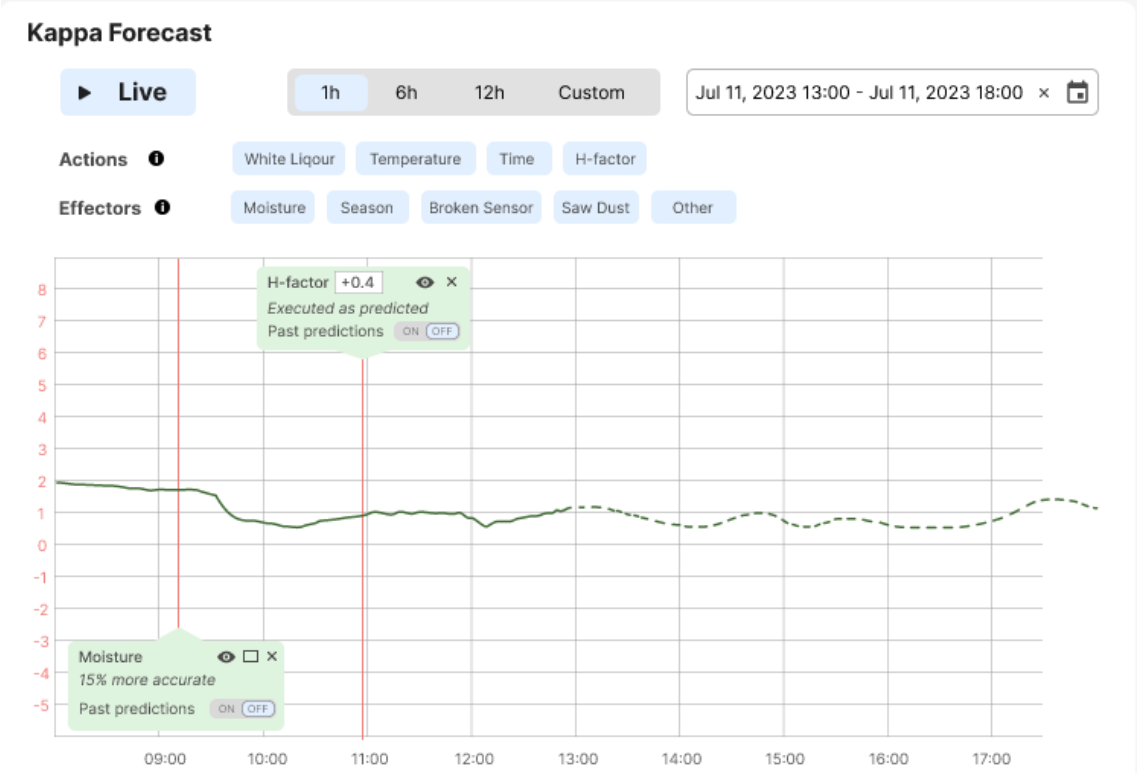


Figure 6.7: The IPG component showing the post-event prediction accuracy when the prediction was accurate.

Figure 6.8: The IPG component showing the post-event prediction accuracy when the prediction was inaccurate.

clusion of this feature can be motivated by the Fogg Behavior Model [102, 103, 104], as easy-to-do actions require a lower threshold of motivation to be acted upon. To enable this three radio buttons were implemented with the option of communicating that the predicted kappa is either "higher than expected", "lower than expected, or "as expected". Only selecting one of these options is the minimum amount of feedback the operators can provide. However, if they want to add additional context to this selection, the design also enables them to drag and drop sensors into this selection, and to add a comment where they can further motivate their selection. Due to the vast amount of unexpected scenarios that can arise during operation, the IET component has to include open-ended feedback features, where the operators can freely explain their reasoning. This part of the IET component enables that. During the evaluations, the operators stressed the importance of displaying previous comments, which is why this feature is based on Version B of the interface. Furthermore, they wished to display more details for the previous comments, such as an exact period for the selected interval and a visualization of the prediction at the time when the comment was created. This has been added to the final design.

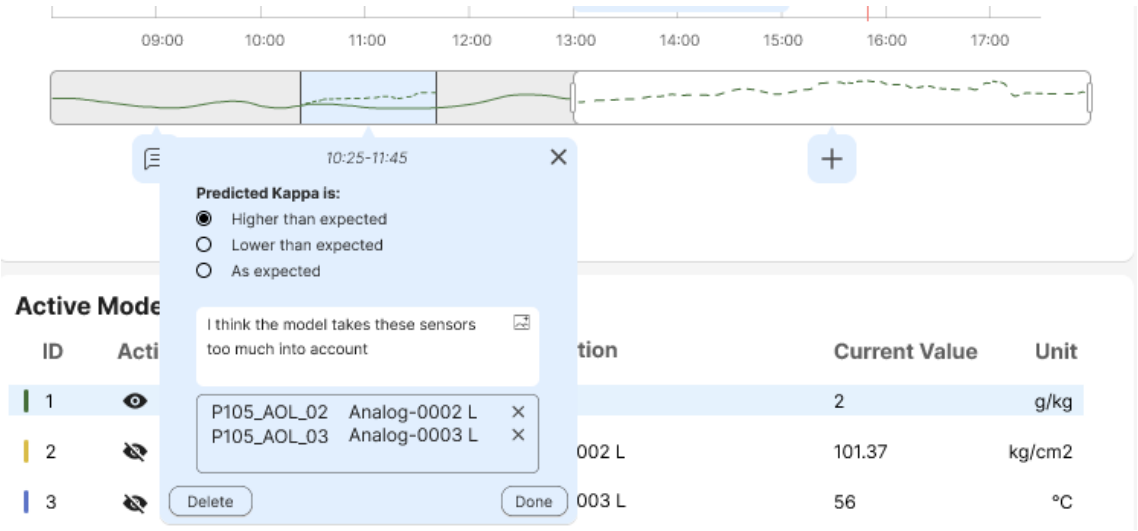


Figure 6.9: The graph navigation tool showing the time span and prediction the comment is referring to, with the predicted kappa selected to be higher than expected, a free text comment, and affected sensors added into the comment.

7
Discussion

In this chapter, the research questions will be answered and discussed in relation to the related literature. We present eight design recommendations for sustaining feedback provision motivation, discuss them with the related literature, and delve deeply into these design recommendations concerning generalizability, showing that they adhere to a set of generalized design guidelines for human-AI interaction [121]. We will also look closely at the importance of UCD approaches for this type of research, examine in detail how the IET components relate to motivation, and through this discussion weave in its relation to the answers to our research questions and design recommendations. The integration of gamification elements will also be discussed, and the methodological limitations of the research conducted in this thesis will be highlighted. Ethical considerations of the research will also be presented and future iterations of the final design will be discussed, laying the foundation for additional research.

7.1 Answering the Research Questions

Previous research shows that there is a need for features similar to the IET components [2, 14, 48], and there are some suggestions on how to implement such features [11, 14, 116]. There is however a gap in research exploring the necessary motivations required for sustained usage of these features. Hence, the research questions chosen for this thesis provide direct value to this missing aspect of human-AI interaction. In this section, the three research questions that were presented in Section 1.1 in Chapter 1 will be presented and answered based on the research conducted.

7.1.1 RQ1

The first research question was *What kinds of motivations are necessary to encourage users to offer feedback to the Explainable Artificial Intelligence system throughout the life cycle of industrial Artificial Intelligence?*
Many of the operators mentioned that to be encouraged to continually offer feedback to the XAI system, they would need to get something in return from the system. This type of response is related to *extrinsic* motivation, as discussed in Section 3.3 in Chapter 3. Getting something in return from the system was mentioned in the initial expert interviews (Section 6.1 in Chapter 6), second participatory design workshop (Section 6.2 in Chapter 6), and evaluations (Section 6.3 in Chapter 6). Based on the initial expert interviews, an important factor to consider is to show that the feedback

improves the AI, which directly corresponds to the operators getting something in return for giving feedback. However, waiting for incremental improvements to the AI model to be implemented can take a long time. Slowly and gradually improving the AI models cannot be solely relied on as a motivator. As demonstrated in the results of the second participatory design workshop (Section 6.2 in Chapter 6), one operator stated that if some types of feedback are only useful for the AI, they are not going to use it in the long run as it needs to be useful for the operators as well. Based on these results, it is clear that the operators desire a direct benefit for giving feedback and interacting with the AI model.
The IPG functionality addresses this. As it simulates, updates, and changes the prediction based on the feedback given by the operators, it also offers insight into future scenarios and into the model itself. Instead of only serving the AI, the IPG functionality serves the operators as well by providing them with an exploratory tool. Simultaneously, the IPG functionality offers rich contextual feedback and data labeling to the AI model, improving it in the long run. When confronted with the IPG functionality in the evaluations (see Section 6.3 in Chapter 6), the operators expressed that it is a feature they think would enable them to be proactive and help them make more educated guesses.
Furthermore, the addition of gamification elements relating to improvement statistics was shown to be motivating to the users and would contribute to the usage of the IET components. This relates to *intrinsic* motivation [57, 59].
In summary, both *extrinsic* motivation and *intrinsic* motivation are necessary to encourage users to offer feedback to the AI system throughout its entire life cycle. Furthermore, what they expect in terms of *extrinsic* motivation is a deeper insight into the AI model, process, and current or future situations, as a response to the feedback they provide. The IPG functionality is one example of how to solve the need for *extrinsic* motivation. Showing improvement statistics was also shown to be motivating to the users, which contributes to *intrinsic* motivation [57, 59].

7.1.2 RQ2

The second research question was *What interface components contribute to motivating the users to offer feedback to the Explainable Artificial Intelligence system?*
Certain interface components and their corresponding interaction modalities contribute to motivation more than others. Firstly, as seen in the results of the evaluations (Section 6.3), the gamification elements related to points and working towards milestones, although motivating in theory, did not contribute to the usage of the IET component. This research showed that the operators do not want to collect points and compete against each other or towards collective milestones. However, what they find motivating are statistics showing how the models have improved based on the feedback they have provided. Secondly, efficient interaction modalities involving as little free text writing as possible were preferred. For example, as seen in the results of the evaluations (Section 6.3 in Chapter 6), the operators appreciated the drag-and-drop functionality between the sensor overview and the IET components in the interface. Having efficient interaction modalities enables the operators to provide rich information to the system in a short amount of time. This is related to

Fogg's Behavior Model [102, 103, 104], as increased ability to perform an action (i.e. more efficient interactions) requires less motivation for the action to be executed.
In summary, improvement statistics are interface components that contribute to motivation, and any interface components offering efficient interaction modalities also contribute to the usage of the IET components by requiring a lower threshold of motivation to be used.

7.1.3 RQ3

The third research question was *What design recommendations can be identified for providing feedback to the Explainable Artificial Intelligence system?*
The following design recommendations have been identified regarding how to provide feedback to an XAI system. These recommendations stem from the results of the initial expert interviews (Section 6.1 in Chapter 6), second participatory design workshop (Section 6.2 in Chapter 6), and evaluations (Section 6.3 in Chapter 6). Naturally, some of these recommendations overlap with the previous research questions. With that in mind, these recommendations can be seen as a summary of all results from the research done in this thesis.

1. **Longevity hinges on immediate usable insights.** To ensure longevity and continued usage of the feedback solution, the act of giving feedback to the system should be immediately insightful for the user. This means that the act of giving feedback should result in an increased understanding of the AI, the related process, and the user's current situation. Obtaining such insight requires an immediate response from the system based on the users' actions. Insight is gained from good explainers, and since the ability to give additional accurate feedback to the system is greatly dependent on this [116], we recommend that explainers are baked into the feedback solution. Furthermore, utilizing counterfactual examples and hypothetical scenarios that explore the decision-making of the AI as a tool for giving feedback to the system is an effective way to address all these points. As demonstrated by the IPG functionality, feedback, insight, and refined predictions can be combined within a single solution. Previous research has also shown that highlighting the relationship between user interaction and predictions influences the users' preferences for a system [117], supporting the idea that counterfactual examples and hypothetical scenarios are sound approaches to ensure longevity and continued usage of the feedback solution.
2. **AI model selection enhances trust and feedback.** Having several AI models to choose from with slightly different parameters, offers quick feedback as the most suitable model gets chosen more often based on accuracy and performance. This ensures a natural selection of AI models, providing insight into what type of model configuration works best. An automatic model selector with the same purpose has already been explored [118], and perhaps this could be done automatically in the future. However, there is related research showing that allowing users to explore contrasting features in predictions plays a role in user trust [119]. By generalizing these insights, allowing the user to select between models could have a positive impact on trust, while simultaneously

serving as a quick and easy-to-do feedback solution.

3. **Quick feedback options promote usage.** An option for giving feedback should be quick and brief, ideally as simple as selection boxes. This can be based on the Fogg Behavior Model [102, 103, 104], as easy-to-do interactions require a lower threshold of motivation to be acted upon. There is a trade-off here, where more brief types of feedback are less informative, but to ensure prolonged usage of the feedback solutions, such feedback should be an option.
4. **Efficient interaction modalities facilitate richer feedback.** The feedback solution should also allow for richer modes of feedback, which should be done through efficient interaction modalities to lower the threshold of motivation required to do so [102, 103, 104]. For example, dragging and dropping larger chunks of information from other parts of the interface into the feedback solution, compared to manually describing that information in text.
5. **Text feedback should be an option.** Although ideally avoided as it is cumbersome, written text feedback offers flexibility when additional context is required [120], and should therefore be an option within the feedback solution.
6. **AI improvement statistics motivate users.** The system should show how feedback is used in a meaningful way by, for example, displaying improvement statistics showing how the AI has improved as a result of feedback. This relates to gamification, as it establishes clear goals and purposes for the act of giving feedback, which can be a reason why it increases motivation in users [59]. This can also be connected to *belonging* in Fogg's Behavior Model [103, 104], which is a social driver of motivation. *Belonging* specifies that humans pursue certain behaviors to improve social status and a sense of belonging.
7. **Avoid scoring systems and rewards.** Scoring systems and rewards as a means of increasing motivation should be avoided unless it is highly unobtrusive and strictly opt-in [57, 63].
8. **Display past feedback.** Previous feedback should be visible in the interface so that users can "compare notes" and learn from each other. Previous feedback should also be coupled with the reasoning and decision-making of the AI at that time. This ensures feedback is coherent between users and that they understand how feedback is given in relation to the reasoning and decision-making of the AI. Furthermore, this can also be connected to *belonging* in Foggs Behavior Model [103, 104], specifying that humans pursue certain behaviors to improve social status and a sense of belonging.

7.2 Summarizing Our Contributions

In this research, we have found that to fully address the question of motivation, the IET components need to be shaped according to motivation from the get-go. Furthermore, the IET components and their corresponding strategies for increasing motivation need to be designed with close adherence to the specific use case, in this case, the delignification process. It could be the case that specific strategies for increasing motivation vary depending on the context, which underlines the importance of UCD approaches. Through the research conducted in this thesis, we have directly addressed the research questions by 1) explaining that both *intrinsic*

and *extrinsic* motivations are required for sustained usage of the IET components, with a corresponding strategy suitable for this particular use case, 2) identified AI improvement statistics as important inclusions in interface components, and that efficient interaction modalities increase the likelihood of using the IET components, and 3) identified eight design recommendations grounded in both previous research as well as the research conducted in this thesis.

7.3 Benefits of a User-Centered Design Approach

The main aim of this research was to understand what kinds of motivations and which interface components contribute to encouraging users to provide feedback to an XAI system. To address this, an interface enabling feedback to an XAI system has been designed with careful consideration of previous research and with close collaboration with end-users. The feedback solutions in the IET components have all been designed according to UCD approaches and have followed the scientific recommendations on Human-Centered XAI and UX Design outlined in Sections 3.2.1 and 3.5 in Chapter 3. All activities in this thesis, such as the background research into this specific use case, the initial expert interviews, the second participatory design workshop, and the evaluations, were chosen because of the close adherence to UCD principles. We argue that this approach is a strong advantage to the research conducted in this thesis and that the results benefit greatly from it, as the insights presented in all research questions are a direct consequence of this approach.

Some of the design recommendations could probably have been arrived at through previous research, but knowing that those design recommendations would be tailored to this specific use case, is less obvious. It is solely through the UCD approach that the results are anchored in this use case. Some design recommendations however, such as **Longevity hinges on immediate usable insights** and **Display past feedback**, are less obvious when looking at previous research. Throughout the UCD activities, nearly all operators stressed the importance of immediate usable insights as a result of providing feedback and that showing past feedback was crucial for both understanding the AI as well as how to provide cohesive feedback. These are important contributions. Furthermore, the answer to the first and second research questions, stating that both *extrinsic* and *intrinsic* motivations are necessary, and that improvement statistics and efficient interaction modalities contribute to motivation, could also have been arrived at through previous research. However, understanding *how* and *why* these aspects contribute to motivation and the usage of the IET components is a direct result of the user-centered research conducted in this thesis.

Auernhammer [15] describes the importance of identifying users' needs early on in the design process. This was done in two ways, by incorporating the previous background research done by ABB into our design, as well as identifying further needs specifically related to feedback in the initial expert interviews. These initial explorations created the foundation to ensure a UCD approach. Another important aspect of our UCD approach is the participatory design workshops.

7.3.1 Participatory Design Workshop with Expert Users

As described in Section 4.4.2 in Chapter 4, participatory design workshops are used to allow the users to be the experts of their own experience [98, 111]. Participatory design is beneficial to ensure that the design adheres to what the users want, how they behave, and that it is fitting to their specific context. These are important considerations in UX design according to the Nielsen Norman Group [76]. Furthermore, participatory design is described as a necessary approach by Auernhammer [15], to ensure that the AI system is usable for the users.
Taking this user-centered approach was extremely valuable as the final design is something created for the operators, by the operators [22, 45]. Also, given the complexity of the technology behind such a system and how the KRAFT process works [6], it was very insightful to have the operators' direct input regarding what is possible and what they want in the interface. For example, the design recommendation **Longevity hinges on immediate usable insights**, is a direct result of this workshop.
As a result of the UCD approach taken in this research, we argue that the operators' contributions, where they directly could discard anything unnecessary in the interface, also contributed to the avoidance of excise in the design [52]. As discussed in Section 3.4.3 in Chapter 3, Cooper et al. [49] suggest that to avoid excise, the number of places a user must navigate should be minimized, which is something the operators directly contributed to. In general, the participatory design workshop involving the operators was a crucial aspect of this research that served to legitimize and ground the resulting design in reality.

7.4 IET Components and Motivation

The final design of the interface and the IET components adhere to what Cooper [51] describes as a sovereign posture interface. In short, the design is highly detailed and requires some initial learning to be used. Once learned, however, the effort required to give feedback is dynamic. The IET components offer both brief easy-to-do feedback, as well as longer and richer methods for giving feedback. These insights are generalized in the design recommendations **Quick feedback options promote usage** and **Efficient interaction modalities facilitate richer feedback**. These differences can affect the usage of the related IET components. Furthermore, the IPG functionality has been demonstrated to be valuable for the operators, which directly increases the motivation for it to be used. In this section, we will highlight how the IET components relate to motivation.

7.4.1 IPG Functionality

As mentioned in Section 3.2.1 in Chapter 3, the linguist Paul Grice [96] outlines important considerations for what constitutes effective communication. At the top of his list is *quality*, meaning that the conveyed information should be of high quality and true. An important consideration for the usage of the IET components is the *quality* of communication between the XAI system and the operators, i.e. the

accuracy of the predictions. The results in this thesis seem to indicate that the motivation to use the IET components is related to the accuracy and precision of the AI models (Sections 6.1.2.5 and 6.1.2.9 in Chapter 6). The purpose of the IET components is to gradually improve the AI over time, as the training data is oftentimes not enough to create an initially accurate AI model. Because of this, there is a risk that the initial *Quality* of information conveyed by the AI is poor. As we have found, if the AI is not useful for the operators, motivation to use it will diminish as the novelty of the system wears off.
To address this, the IPG functionality was created. According to the research in this thesis, the IPG functionality has the potential to become a valuable tool for operators, while simultaneously providing feedback for the AI. Even if the AI-produced predictions are not fully accurate in the beginning, neither are the operators' predictions as seen in Section 6.1.2.3 in Chapter 6. As seen in the results of the initial expert interviews (Section 6.1 in Chapter 6), the operators rely on communication and cooperation with each other to figure out the best strategy going forward. The IPG functionality offers such communication and cooperation between the AI and the operators; it essentially becomes a partner to bounce ideas off of. Counterfactual questions and explanations, as offered by the IPG functionality, are crucial to understanding the AI, avoiding bias, as well as fostering trust in its decisions [3, 34]. As shown in the results of the initial expert interviews (Section 6.1 in Chapter 6), these attributes are necessary building blocks to ensure motivation to give feedback. Furthermore, the results of the evaluation indicate that the IPG functionality does indeed have the potential to increase understanding and trust (see Section 6.3 in Chapter 6). These results are generalized in the design recommendation **Longevity hinges on immediate usable insights**.
In short, the motivation to use the IPG functionality is mainly based on the value it provides for the operators. So far, this thesis has shown promising results regarding this functionality, but to be successful, the implementation of it also has to live up to the same promises.

7.4.2 Model Selection And Dynamic Comments

According to Fogg's Behavior Model, three things are needed for a behavior to happen: *motivation*, *ability*, and *trigger* [103, 104]. Users require less motivation to perform a certain action if it is easy to do. Since one of the core parts of this research has been about ensuring long-term usage of the IET components, we have used the Fogg Behavior Model as a basis for understanding how some of the interactions in our interface contribute to increased *ability*. One missing aspect of the model in our design is the lack of explicit *triggers*. However, we reason that a wrongful prediction would be enough to trigger the users into giving feedback to the system. In that case, to promote the usage of the IET components, it was necessary to ensure that the interactions were simple enough to require a low threshold of motivation to be used. This is specified in the design recommendations **AI model selection to enhance trust and feedback**, **Quick feedback options promote usage**, and **Efficient interaction modalities facilitate richer feedback**.
The model selection in the interface (Fig. 6.5) is one such component that gives the

easiest form of feedback to the system. The selection of models is in itself a mode of feedback that the data scientists can look at to see which models are the most used. This is therefore a suitable action for the operators with low motivation as it allows them to quickly provide feedback with low effort.
The dynamic comments component (Fig. 6.9), on the other hand, could require more time and effort from the operators, depending on how much information the operators want to convey. As Fogg states, the less time it takes to perform an action, the more likely it is for the user to perform that action. The reason behind the name *dynamic* comments is that the operators do not *have* to fill everything in. They could choose to, for example, only select if the predicted kappa is higher, lower, or as expected and submit the comment. This makes the threshold for giving feedback to the system low, while still being valuable to the data scientists and the improvement of the AI models. Giving the quickest form of feedback within the comment means that it requires less mental effort from the operators which also increases the likelihood that they use the functionality more. Again, the second participatory workshop was useful in this regard, since the operators were clear about their preference for simplifying the quickest form of feedback in the dynamic comments. This shaped the design to change from sliders to radio buttons to communicate their level of agreement toward a prediction. This also resulted in the design recommendation **Quick feedback options promote usage**.
Also related to the interface as a whole but specifically the dynamic comments is the concept of interaction cost, as described in Section 3.5.1 in Chapter 3. Similarly to the mental effort described by Fogg, interaction cost is the cognitive effort of a design [77, 78]. To ensure that the operators have everything they need to produce a comment, we have tried to lower the interaction cost in the interface by keeping everything related to the comments nearby and thus reducing the need for navigating to different parts of the UI [79]. As seen in the evaluations, being able to quickly view previous comments was an appreciated feature that enabled operators to learn from each other and create coherent comments. The design recommendation **Display past feedback** was a result of this. Furthermore, the ability to drag and drop sensor information into the comments effectively reduces the interaction cost, as the operators do not have to remember and manually input that information into the comment.

7.4.3 Gamification and Motivation

One common way to facilitate motivation is to use gamification, as was discussed in Section 3.3.2 in Chapter 3.
Research that had been done on gamification specifically for Industry 4.0 [60, 57] showed that it is motivating because it establishes clear goals and purposes for the tasks. In the evaluations, it was found that the operators did not want to collect points, compete against each other, or work toward collective milestones. On the contrary, they found it to be distracting from their actual work. This insight is generalized in the design recommendation **Avoid scoring systems and rewards**. What did motivate them, however, was to see that the models had improved, which shows that their feedback helped the model progress. Improvement statistics such

as these relate to some aspects of gamification as described by research [62, 59], as it does establish a clear goal and purpose for the act of giving feedback. This is also clarified in the design recommendation **AI improvement statistics to motivate users**, as well as in the answer to the second research question.
However, the opinions regarding the other types of gamification, such as scoring systems and working towards milestones, were divided. In the second participatory design workshop, the operators were generally positive about the idea, as long as it was made in such a way that they were working towards common goals and milestones and not competing against each other. One reason for the discrepancy of opinions could be that the version presented in the second participatory design workshop was only loosely defined and that the implementation created for the evaluation did not completely adhere to the research in the area. For example, research by Reis et al. [57] suggests that scoring systems could have a negative impact on motivation as it could become a primary focus for the users. The design made for the evaluation included such a scoring system, and the negative impressions of it strengthen the reasoning of Reis et al. [57]. Furthermore, the scoring system presented in the evaluations where not designed as voluntary. The users simply gathered points while interacting with the system whether they wanted to or not. Again, research on gamification suggests that the usage of any gamification techniques should be voluntary [63].
The only aspect of gamification that all operators liked was the improvement statistics, showing that their feedback contributes to better models over time. This establishes a clear goal and purpose for the act of giving feedback. Had the other gamification techniques been implemented in a way more closely adhering to the research in the area, perhaps a different result would have emerged.

7.5 Generalizability of the Design Recommendations

The entire thesis used a research-through-design approach, meaning that the results presented here apply to the context in which they were researched and designed for. However, we believe that the insights from this study could be applicable in other contexts as well. Regarding the generalizability of the results, the design itself is specific to this use case, however, the design recommendations and the understanding of the types of motivation required for prolonged use of the IET components could be more universal. For example, the design recommendations that emerged as a result of this research can be closely related to previous research in the area of Human-AI interaction by Amershi et al. [121] who provides 18 design guidelines for Human-AI interaction.
Firstly, Amershi et al. [121] underline the importance of enabling users to give granular feedback indicating their preferences while interacting with the system. This corresponds with design recommendations **AI model selection to enhance trust and feedback** and **Quick feedback options promote usage**, which state that brief, easy-to-do feedback is necessary for feedback functionalities to help promote usage over long periods.

Secondly, Amershi et al. [121] state that the AI system should convey the consequences of user actions, and immediately update or convey how user actions will impact future behaviors of the AI system. This closely corresponds to design recommendation **Longevity hinges on immediate usable insights**, which underlines the necessity of an immediate insightful response based on user actions and their given feedback. Additionally, they also state that the interface should support efficient correction and make it easy to edit, refine, or recover when the AI system is wrong. This corresponds to the IPG functionality, which offers an easy method for refining and improving the prediction, offering feedback to the system, while still being insightful for the user.
Thirdly, Amershi et al. [121] state that the system should remember recent interactions, maintain short-term memory, and allow the user to make efficient references to that memory, which corresponds to design recommendation **Display past feedback**, clarifying that previous feedback should be visible in the interface together with the AI's prediction at that time.
These similarities provide some insight into the generalizability of our design recommendations even though they are the result of a highly specific use case with highly specified expert users. Of course, additional research is required to fully understand the generalizability of the design recommendations, but based on these connections to similar research, the argument of generalizability can be strengthened.

7.6 Limitations

This section will cover limitations, alternative approaches, and arguments supporting the legitimacy of the approaches used in this thesis.

7.6.1 Bleaching Operators

An obvious limitation throughout this research is the participation of mainly bleaching operators instead of operators directly working with the delignification process. Although they have very similar working conditions, there are slight differences in the process they work with. For example, the bleaching process is less sensitive to variance of incoming product from the previous steps in the process since the bleaching of the mass can be compensated for at the very end. The delignification process requires more careful consideration of actions since changes to the process can have unintended consequences four to six hours down the line which cannot be as easily compensated for. However, several of the bleaching operators had significant experience working in the delignification process as well. Especially the three participants in the evaluations, two of whom had over eleven years of working experience in the factory and one who had seven to ten years, came with a lot of knowledge of every department in the factory. They all displayed a great understanding of the delignification process. In retrospect, the operators' experience of the delignification process should have been reported as a part of the demographics for each operator, this would have ensured more transparent results. However, due to the overwhelming amount of similarities between being an operator in the bleaching process and being one in the delignification process, we argue that the results are generalizable

to the delignification process. Also, since the context in which the operators work is quite niche - sitting in a control room and moderating the paper production process - we argue that the bleaching operators were the closest to the delignification operators and their insights were still incredibly valuable for the research.
Furthermore, the bleaching operators had direct experience with an AI with a feedback functionality deployed in their workplace. Since this AI was poorly received, interviewing these operators gave us an in-depth understanding of what to avoid and why. This insight is a direct benefit of interviewing these particular operators.

7.6.2 Limited Amount of Expert Users in the Evaluation

Another point of discussion is the number of expert users in the evaluations. While there were only three expert users in the evaluations, a total of 12 unique operators contributed to the design through the expert interviews, the second participatory design workshop, and evaluations. Four UX designers also contributed to the design through the first participatory design workshop. This means that the entire research had 16 unique participants which we argue is enough to reach a valid conclusion. Considering that 12 of these participants are expert users, this number is in line with previous research conducted with expert users [106, 108, 109].
For evaluations, some research suggests that three participants are enough, as the first three participants usually encounter most of the usability issues in the design and when the number of participants increases, recurring patterns are revealed [113]. Since we started seeing recurring patterns in the responses and actions of the participants already at only three expert users, a point of satiation was arguably reached. Furthermore, since the design was directly developed together with the operators during the second participatory design workshop, the most major usability issues likely never surfaced as the design was already approved by the operators.

7.6.3 Summative Interview Questions

Most commonly, it is suggested to avoid using yes or no questions while interviewing users as it could affect the validity of the obtained information [126]. This is something we should have considered regarding the summative interview questions as they predominantly could have been answered using yes or no. However, since the interview data was analyzed together with some of the insights from the think-aloud session, we had access to their raw impressions of the IET components, not affected by the poor wording of the questions. We argue that this strengthens the validity of the results. Also, even if the operators responded with a yes or no to the questions, they often elaborated on why they answered yes or no.

7.6.4 Potential Bias in the Second Participatory Design Workshop

In the second participatory design workshop, as was described in Section 5.3.5 in Chapter 5, the participants were given cut-out paper wireframes to freely place, add, or remove elements from. Since the operators were given these pre-designed

wireframes, one limitation of the session could be that they were primed by the designs and therefore less inclined to think of their own design suggestions. On the other hand, given that the operators are expert users and not designers, we argue that we get better results by providing them with existing design solutions rather than having them design something from scratch, as that can be intimidating for someone not used to the practice of sketching and designing [98]. Despite this, many operators decided to sketch on their own and added or removed things from the designs. Having pre-designed wireframes also allowed us to gain insight into their thoughts and opinions of where the design was headed, which worked as a sort of formative evaluation guiding the future directions of the design.

7.7 Ethical Considerations

The process of interviewing, conducting design workshops, and evaluating involved addressing participants' privacy concerns and ensuring that their information was handled with care and respect during and after the interaction. During all activities with participants, we provided them with consent forms, exercised caution in handling personal data, prioritized privacy, and complied with ethical standards and regulations throughout the process such as the General Data Protection Regulation (GDPR). All collected data throughout this thesis has been anonymized to ensure the privacy of the participants. The development of solutions for deployed XAI systems required a keen focus on safety, emphasizing clear communication between humans and machines to mitigate potential hazards in the production process where the XAI systems are used. Additionally, due to the confidential nature of the company and its partnerships, strict adherence to nondisclosure agreements (NDAs) was paramount. This obligation necessitated careful handling of information, limiting the sharing of details about the work during the thesis writing process.

7.8 Future Work

This research has aimed to create feedback features that can be used to improve the accuracy of AI models, however, future work should also address how similar feedback features can be used to improve the explainability of the XAI interface. Currently, the IPG functionality contributes to increased explainability, allowing the predictions to become more transparent. Largely, however, the area remains unexplored in this research as the primary focus was on increasing motivation for feedback provision.
Since XAI and giving feedback to AI systems are still fairly unexplored areas, there are many possibilities for future research. With regards to feedback and motivation specifically, it would first and foremost be interesting to conduct more research similar to what has been done in this thesis, to further understand users' needs and perceptions of giving feedback to an XAI system in this way. Involving more than 12 expert users would be a good suggestion for future research, as it would yield broader and deeper insights. Considering the results of this thesis, an obvious suggestion for future research is to implement the functionalities described in our design and

conduct longitudinal studies examining the usage and user perceptions of the IET components. It would also be interesting to test several more aspects and different implementations of gamification since previous research shows promise for increasing motivation and acceptance in industrial contexts [57, 59]. With implementations of gamification more closely adhering to the suggestions of previous research [58, 60, 61, 62, 63], perhaps a different outcome could be reached and gamification could successfully contribute more to motivation to use the IET components.
After finalizing this research, some additional insights into the design of the IPG functionality came to mind. However, due to a lack of time and insufficient user input justifying any major design changes, we did not implement these. For example, the operators have mentioned in the second participatory design workshop and the evaluations that being able to mark the period where an *action* or an *effector* affects the process is a feature they would like to have. Currently, they simply type the duration into the interface, but there is no visual communication indicating that duration. Taking inspiration from video editing software could be an option, as shown in Fig. 7.1. Here the duration of the *action* or *effector* is visually communicated, and by dragging the box (similar to Version A of the interface, Fig. 5.10), the operators can increase or decrease the duration. This alternative suggests how the IPG component's design could look, but more explorations and testing would be required to find the best possible design solution.
While this thesis researched a very specific context - the delignification process - an XAI system giving predictions and the users giving feedback on the predictions is something that could be useful in other contexts as well. It could be useful in other production processes, not only the paper production process. XAI could also be used in various decision-making processes, for example in the medical field [122]. If a similar prediction system were to be used in a medical context, users' feedback and interaction with the system would be crucial to ensure accurate detection, diagnosis, treatment, outcome predictions, and prognosis evaluations [123]. In such cases, some of the design recommendations found through the research in this thesis could be generalized to these contexts as well.

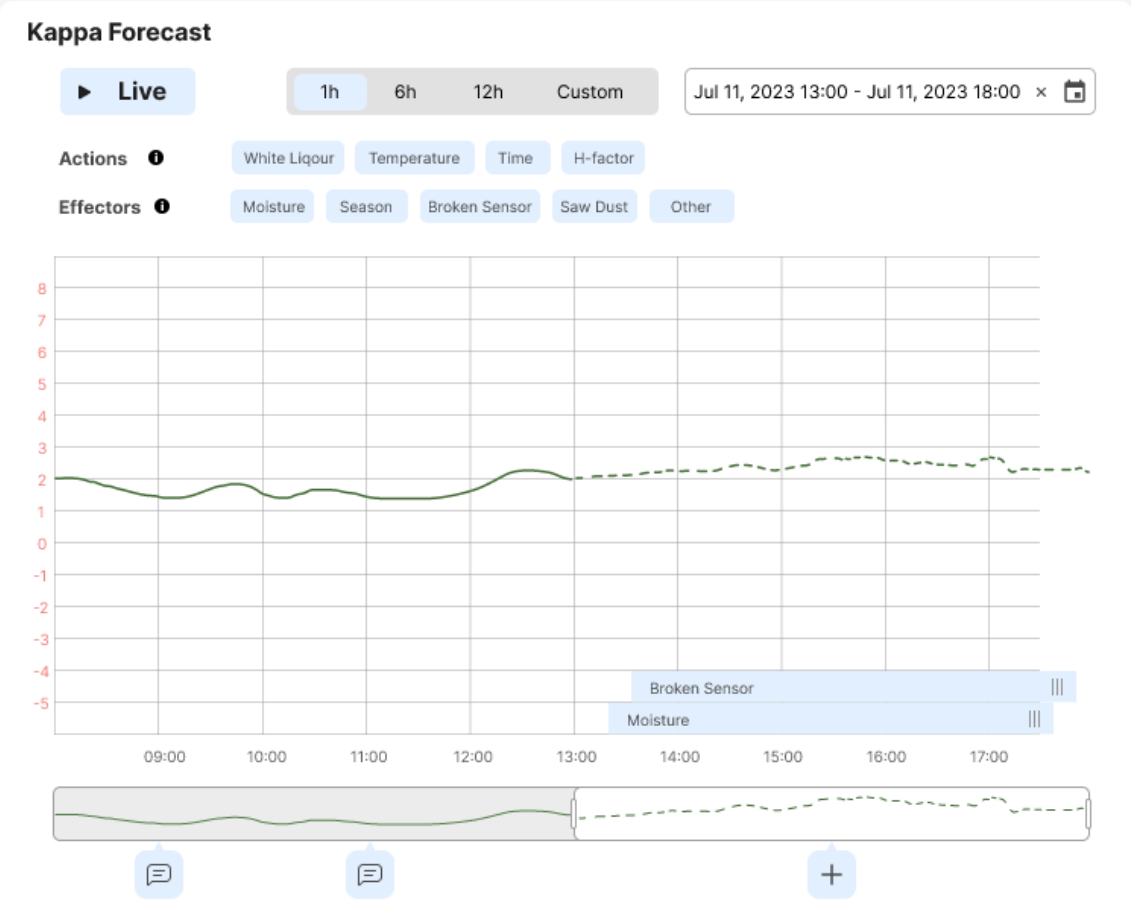


Figure 7.1: A suggestion for the future direction of the IPG where the added action and effector are represented by longer boxes.

8

Conclusion

The purpose of this thesis was to explore what kinds of motivation are necessary to encourage users to give feedback to an explainable artificial intelligence (XAI) system used to give predictions in an industrial process, what interface components contribute to motivating users to give feedback to the XAI system and to identify design recommendations for giving feedback to XAI systems.

By following one iteration of the double diamond design process, using a user-centered approach with methods such as expert interviews, brainstorming, participatory design workshops, rapid prototyping, and evaluations, the thesis presented suggestions for the design of Incremental Explanatory Training (IET) components for interacting with and giving feedback to predictions made by the AI. The research questions are directly addressed by 1) explaining that both *intrinsic* and *extrinsic* motivations are required for sustained usage of the IET components, with a corresponding strategy suitable for this particular use case, 2) identifying AI improvement statistics as important inclusions in interface components, and that efficient interaction modalities increase the likelihood of using the IET components, and 3) identifying eight design recommendations grounded in both previous research as well as the research conducted in this thesis.

The design solution included an Interactive Prediction Graph (IPG) that allows the operators to explore "what-if" scenarios with the prediction by placing hypothetical actions and variables that might affect the process directly into the graph. The addition of these hypothetical actions and variables updates and refines the prediction, provides context for it, and labels the data so that the AI can iteratively be improved by data scientists. Simultaneously, it is a valuable tool for the operators allowing them to collaborate with the AI. This contributes to the operators' *extrinsic* motivation to use the IET components, as they get a deeper insight into the reasoning of the AI, the production process, and future and current situations, as a response to the feedback they have provided.

Furthermore, gamification was explored as a technique for motivating the operators. More precisely, the results show that interface components including gamification elements such as improvement statistics of the AI, showing how it has improved as a consequence of feedback, could support *intrinsic* motivation to provide feedback. Ease of use was also demonstrated to be a critical factor relating to motivation. The interface components offering efficient interaction modalities were preferred, showing that if the feedback is to be used long-term, it needs to be as frictionless as possible. Eight generalized design recommendations for designing feedback for AI systems could also be identified, as shown in Table 8.1. This thesis has laid the foundation for future research aimed at optimizing user-provided feedback to XAI systems.

Number	Recommendation	Description
1	Longevity hinges on immediate usable insights.	To ensure longevity and continued usage of the feedback solution, the act of giving feedback to the system should be immediately insightful for the user. This means that the act of giving feedback should result in an increased understanding of the AI, the related process, and the user's current situation.
2	AI model selection enhance trust and feedback.	Having several AI models to choose from with slightly different parameters, offers quick feedback as the most suitable model gets chosen more often based on accuracy and performance. This ensures a natural selection of AI models, providing insight into what type of model configuration works best.
3	Quick feedback options promote usage.	An option for giving feedback should be quick and brief, ideally as simple as selection boxes, as easy-to-do interactions require a lower threshold of motivation to be acted upon.
4	Efficient interaction modalities facilitate richer feedback.	The feedback solution should also allow for richer modes of feedback, which should be done through efficient interaction modalities to lower the threshold of motivation required for usage.
5	Text feedback should be an option.	Although ideally avoided as it is cumbersome, written text feedback offers flexibility when additional context is required.
6	AI improvement statistics motivate users.	The system should show how feedback is used in a meaningful way by, for example, displaying improvement statistics showing how the AI has improved as a result of feedback.
7	Avoid scoring systems and rewards.	Scoring systems and rewards as a means of increasing motivation should be avoided unless it is highly unobtrusive and strictly opt-in.
8	Display past feedback.	Previous feedback should be visible in the interface so that users can "compare notes" and learn from each other through previous interactions with the AI.

Table 8.1: Summary of the design recommendations created in this thesis.

Bibliography

[1] Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., ... & Baum, K. (2021). What do we want from Explainable Artificial Intelligence (XAI)?–A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. Artificial Intelligence, 296, 103473.

[2] Liao, Q. V., Gruen, D., & Miller, S. (2020, April). Questioning the AI: informing design practices for explainable AI user experiences. In Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1-15).

[3] Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019, May). Designing theory-driven user-centric explainable AI. In Proceedings of the 2019 CHI conference on human factors in computing systems (pp. 1-15).

[4] Woodie, A. (2018, May 30). Opening Up Black Boxes with Explainable AI. Datanami.com. https://www.datanami.com/2018/05/30/opening-up-black-boxes-with-explainable-ai/

[5] Explain Project (2023). https://explain-project.eu/

[6] Sixta, H., & Schild, G. (2009). A new generation kraft process. Lenzinger Berichte, 87(1), (pp. 26-37).

[7] Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., ... & Zimmermann, T. (2019, May). Software engineering for machine learning: A case study. In 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP) (pp. 291-300). IEEE.

[8] Vogelsang, A., & Borg, M. (2019, September). Requirements engineering for machine learning: Perspectives from data scientists. In 2019 IEEE 27th International Requirements Engineering Conference Workshops (REW) (pp. 245-251). IEEE.

[9] Nazir, S., Colombo, S., & Manca, D. (2012). The role of situation awareness for the operators of process industry. Chemical Engineering Transactions, 26.

[10] Ehsan, U., Wintersberger, P., Liao, Q. V., Watkins, E. A., Manger, C., Daumé III, H., ... & Riedl, M. O. (2022, April). Human-Centered Explainable AI (HCXAI): beyond opening the black-box of AI. In CHI conference on human factors in computing systems extended abstracts (pp. 1-7).

[11] Spinner, T., Schlegel, U., Schäfer, H., & El-Assady, M. (2019). explAIner: A visual analytics framework for interactive and explainable machine learning. IEEE transactions on visualization and computer graphics, 26(1), 1064-1074.

[12] S. Chen, Cai, H., & Zhuang, Y. (2020). XAIson: A Visual Analytics Framework for Interactive and Explainable AI. IEEE Transactions on Visualization and Computer Graphics, 26(1), 9-19.

[13] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

[14] Teso, S., & Kersting, K. (2019, January). Explanatory interactive machine learning. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (pp. 239-245).

[15] Auernhammer, J. (2020). Human-centered AI: The role of Human-centered Design Research in the development of AI.

[16] Interaction Design Foundation - IxDF. (2016, May 25). What is Design Thinking (DT)?. Interaction Design Foundation - IxDF. https://www.interaction-design.org/literature/topics/design-thinking

[17] Schoonderwoerd, T. A., Jorritsma, W., Neerincx, M. A., & Van Den Bosch, K. (2021). Human-centered XAI: Developing design patterns for explanations of clinical decision support systems. International Journal of Human-Computer Studies, 154, 102684.

[18] Zytek, A., Liu, D., Vaithianathan, R., & Veeramachaneni, K. (2021). Sibyl: Understanding and addressing the usability challenges of machine learning in high-stakes decision making. IEEE Transactions on Visualization and Computer Graphics, 28(1), 1161-1171.

[19] Kovalerchuk, B., Ahmad, M. A., & Teredesai, A. (2021). Survey of explainable machine learning with visual and granular methods beyond quasi-explanations. Interpretable artificial intelligence: A perspective of granular computing, 217-267.

[20] Zhao, X., Wu, Y., Lee, D. L., & Cui, W. (2018). iforest: Interpreting random forests via visual analytics. IEEE transactions on visualization and computer graphics, 25(1), 407-416.

[21] Chanda, A. K., Egleston, B. L., Bai, T., & Vucetic, S. (2022, October). MedCV: An interactive visualization system for patient cohort identification from medical claim data. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management (pp. 4828-4832).

[22] Doroftei, D., De Cubber, G., Wagemans, R., Matos, A., Silva, E., Lobo, V., ... & Serrano, D. (2017). User-centered design. Search and rescue robotics. From theory to practice. IntechOpen, London, 19-36.

[23] Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. Contemporary educational psychology, 25(1), 54-67.

[24] Springer, A., & Whittaker, S. (2019, March). Progressive disclosure: empirically motivated approaches to designing effective transparency. In Proceedings of the 24th international conference on intelligent user interfaces (pp. 107-120).

[25] Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

[26] Rahman, M., Avelin, A., & Kyprianidis, K. (2020). A review on the modeling, control and diagnostics of continuous pulp digesters. Processes, 8(10), 1231.

[27] Rahman, M., Avelin, A., & Kyprianidis, K. (2019). An approach for feedforward model predictive control of continuous pulp digesters. Processes, 7(9), 602.

[28] Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE access, 6, 52138-52160.

[29] Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). Metrics for explainable AI: Challenges and prospects. arXiv preprint arXiv:1812.04608.

[30] Amiri, S. S., Weber, R. O., Goel, P., Brooks, O., Gandley, A., Kitchell, B., & Zehm, A. (2020). Data representing ground-truth explanations to evaluate xai methods. arXiv preprint arXiv:2011.09892.

[31] Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K. R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. Proceedings of the IEEE, 109(3), 247-278.

[32] Sharma, S., & Seth, U. (2017). Artificial intelligence in cardiology. Journal of the Practice of Cardiovascular Sciences, 3(3), 158-159.

[33] Rubin, V. (2020, November). AI Opaqueness: What Makes AI Systems More Transparent?. In Proceedings of the Annual Conference of CAIS/Actes du congrès annuel de l'ACSI.

[34] Guidotti, R., Monreale, A., Giannotti, F., Pedreschi, D., Ruggieri, S., & Turini, F. (2019). Factual and counterfactual explanations for black box decision making. IEEE Intelligent Systems, 34(6), 14-23.

[35] Ehsan, U., Wintersberger, P., Liao, Q. V., Mara, M., Streit, M., Wachter, S., ... & Riedl, M. O. (2021, May). Operationalizing human-centered perspectives in explainable AI. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-6).

[36] Letzgus, S., Wagner, P., Lederer, J., Samek, W., Müller, K. R., & Montavon, G. (2022). Toward explainable artificial intelligence for regression models: A methodological perspective. IEEE Signal Processing Magazine, 39(4), 40-58.

[37] Rycroft-Malone, J. (2014). From knowing to doing—from the academy to practice Comment on "The many meanings of evidence: implications for the translational science agenda in healthcare". International journal of health policy and management, 2(1), 45.

[38] Wolf, C. T. (2019, March). Explainability scenarios: towards scenario-based XAI design. In Proceedings of the 24th International Conference on Intelligent User Interfaces (pp. 252-257).

[39] Liao, Q. V., Pribić, M., Han, J., Miller, S., & Sow, D. (2021). Question-driven design process for explainable AI user experiences. arXiv preprint arXiv:2104.03483.

[40] Gunning, D., & Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. AI magazine, 40(2), 44-58.

[41] Mueller, S. T., Veinott, E. S., Hoffman, R. R., Klein, G., Alam, L., Mamun, T., & Clancey, W. J. (2021). Principles of explanation in human-AI systems. arXiv preprint arXiv:2102.04972.

[42] Dhanorkar, S., Wolf, C. T., Qian, K., Xu, A., Popa, L., & Li, Y. (2021, June). 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).

[43] Longo, L., Goebel, R., Lecue, F., Kieseberg, P., & Holzinger, A. (2020, August). Explainable artificial intelligence: Concepts, applications, research challenges and visions. In International cross-domain conference for machine learning and knowledge extraction (pp. 1-16). Cham: Springer International Publishing.

[44] Young, S. (2010). Cognitive user interfaces. IEEE Signal Processing Magazine, 27(3), 128-140.

[45] Mhlanga, D. (2022). Human-centered artificial intelligence: the superlative approach to achieve sustainable development goals in the fourth industrial revolution. Sustainability, 14(13), 7804.

[46] Belle, V., & Papantonis, I. (2021). Principles and practice of explainable machine learning. Frontiers in big Data, 39.

[47] Javaid, M., Haleem, A., Singh, R. P., & Suman, R. (2022). Artificial intelligence applications for industry 4.0: A literature-based study. Journal of Industrial Integration and Management, 7(01), 83-111.

[48] Zhou, J., Gandomi, A. H., Chen, F., & Holzinger, A. (2021). Evaluating the quality of machine learning explanations: A survey on methods and metrics. Electronics, 10(5), 593.

[49] Cooper, A., Reimann, R., Cronin, D., & Noessel, C. (2014). About face: the essentials of interaction design. John Wiley & Sons.

[50] Cooper, A. (2014). Designing for people. In A. Cooper, R. Reimann, D. Cronin, & C. Noessel (Eds.), About Face 4: The Essentials of Interaction Design. John Wiley & Sons.

[51] Cooper, A. (2014). Platform and posture. In A. Cooper, R. Reimann, D. Cronin, & C. Noessel (Eds.), About Face 4: The Essentials of Interaction Design. John Wiley & Sons.

[52] Cooper, A. (2014). Reducing work and eliminating excise. In A. Cooper, R. Reimann, D. Cronin, & C. Noessel (Eds.), About Face 4: The Essentials of Interaction Design. John Wiley & Sons.

[53] Darejeh, A., & Salim, S. S. (2016). Gamification solutions to enhance software user engagement—a systematic review. International Journal of Human-Computer Interaction, 32(8), 613-642.

[54] Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design elements to gamefulness: defining" gamification". In Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments (pp. 9-15).

[55] Hamari, J., & Eranti, V. (2011, September). Framework for Designing and Evaluating Game Achievements. In Digra conference (Vol. 10, No. 1.224, p. 9966).

[56] Xu, Y. (2012). Literature review on web application gamification and analytics. CSDL Technical.

[57] Reis, A. C. B., Silva Júnior, E., Gewehr, B. B., & Torres, M. H. (2020). Prospects for using gamification in Industry 4.0. Production, 30, e20190094.

[58] Cameron, J., & Pierce, W. D. (1994). Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis. Review of Educational Research, 64(3), 363-423. https://doi.org/10.3102/00346543064003363

[59] Lee, J., Kim, J., Seo, K., Roh, S., Jung, C., Lee, H., ... & Ryu, H. (2016). A case study in an automotive assembly line: exploring the design framework for manufacturing gamification. In Advances in Ergonomics of Manufacturing: Managing the Enterprise of the Future: Proceedings of the AHFE 2016 International Conference on Human Aspects of Advanced Manufacturing, July 27-31,

2016, Walt Disney World®, Florida, USA (pp. 305-317). Springer International Publishing.

[60] Schuldt, J., & Friedemann, S. (2017, April). The challenges of gamification in the age of Industry 4.0: Focusing on man in future machine-driven working environments. In 2017 IEEE Global Engineering Education Conference (EDUCON) (pp. 1622-1630). IEEE.

[61] Korn, O., Muschick, P., & Schmidt, A. (2017). Gamification of production? A study on the acceptance of gamified work processes in the automotive industry. In Advances in Affective and Pleasurable Design: Proceedings of the AHFE 2016 International Conference on Affective and Pleasurable Design, July 27-31, 2016, Walt Disney World®, Florida, USA (pp. 433-445). Springer International Publishing.

[62] Baalsrud Hauge, J. M., Stănescu, I. A., Carvalho, M. B., Stefan, A., Banica, M., & Lim, T. (2015, August). Integrating gamification in mechanical engineering systems to support knowledge processes. In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (Vol. 57052, p. V01BT02A015). American Society of Mechanical Engineers.

[63] Liu, M., Huang, Y., & Zhang, D. (2018). Gamification's impact on manufacturing: Enhancing job motivation, satisfaction and operational performance with smartphone-based gamified job design. Human Factors and Ergonomics in Manufacturing & Service Industries, 28(1), 38-51.

[64] Lasi, H., Fettke, P., Kemper, H. G., Feld, T., & Hoffmann, M. (2014). Industry 4.0. Business & information systems engineering, 6, 239-242.

[65] Saad, E., Elekyaby, M. S., Ali, E. O., & Hassan, S. F. A. E. (2020). Double diamond strategy saves time of the design process. International Design Journal, 10(3), 211-222.

[66] Kochanowska, M., & Gagliardi, W. R. (2022). The double diamond model: In pursuit of simplicity and flexibility. Perspectives on Design II: Research, Education and Practice, 19-32.

[67] Knopf, J. W. (2006). Doing a literature review. PS: Political Science & Politics, 39(1), 127-132.

[68] Beyers, J., Braun, C., Marshall, D., & De Bruycker, I. (2014). Let's talk! On the practice and method of interviewing policy experts. Interest Groups & Advocacy, 3, 174-187.

[69] Bogner, A., Littig, B., & Menz, W. (2009). Introduction: Expert interviews—An introduction to a new methodological debate. In Interviewing experts (pp. 1-13). London: Palgrave Macmillan UK.

[70] Al-Samarraie, H., & Hurmuzan, S. (2018). A review of brainstorming techniques in higher education. Thinking Skills and creativity, 27, 78-91.

[71] Wilson, J., & Rosenberg, D. (1988). Rapid prototyping for user interface design. In Handbook of human-computer interaction (pp. 859-875). North-Holland.

[72] Cabrio, A., Hashmati, N., Rabia, P., Tumma, L., Wärnberg, H., Hendriks, S., & Obaid, M. (2023, March). HighLight: Towards an Ambient Robotic Table as a Social Enabler. In Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (pp. 146-150).

[73] Herman, H. (2024, January 10). What is brainwriting? The 6-3-5 technique, explained. Zapier.com. Retrieved February 5, 2024, from https://zapier.com/blog/brainwriting/

[74] Figma. Retrieved February 5, 2024, from https://www.figma.com/

[75] Holzmann, C., & Hutflesz, P. (2014, December). Multivariate testing of native mobile applications. In Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia (pp. 85-94).

[76] Neilsen Norman Group. Retrieved February 5, 2024, from https://www.nngroup.com/

[77] Neilsen Norman Group. Retrieved February 5, 2024, from https://www.nngroup.com/articles/interaction-cost-definition/

[78] Lam, H. (2008). A framework of interaction costs in information visualization. IEEE transactions on visualization and computer graphics, 14(6), 1149-1156.

[79] Hui, B., Gustafson, S., Irani, P., & Boutilier, C. (2008, May). The need for an interaction cost model in adaptive interfaces. In Proceedings of the working conference on Advanced visual interfaces (pp. 458-461).

[80] Neilsen Norman Group. Retrieved February 5, 2024, from https://www.nngroup.com/articles/mental-models/

[81] Sheridan, H., Murphy, E., & O'Sullivan, D. (2023, July). Exploring Mental Models for Explainable Artificial Intelligence: Engaging Cross-disciplinary Teams Using a Design Thinking Approach. In International Conference on Human-Computer Interaction (pp. 337-354). Cham: Springer Nature Switzerland.

[82] National Research Council. 1987. Mental Models in Human-Computer Interaction: Research Issues About What the User of Software Knows. Washington, DC: The National Academies Press. https://doi.org/10.17226/790.

[83] Rook, F. W., & Donnell, M. L. (1993). Human cognition and the expert system interface: Mental models and inference explanations. IEEE Transactions on Systems, Man, and Cybernetics, 23(6), 1649-1661.

[84] Jørgensen, A. H. (1990). Thinking-aloud in user interface design: a method promoting cognitive ergonomics. Ergonomics, 33(4), 501-507.

[85] Neilsen Norman Group. Retrieved February 5, 2024, from https://www.nngroup.com/articles/ten-usability-heuristics/

[86] FICO. Retrieved February 8, 2024, from https://www.fico.com/

[87] Goldschmidt, G. (2016). Linkographic evidence for concurrent divergent and convergent thinking in creative design. Creativity research journal, 28(2), 115-122.

[88] Fernandes, Í., Rocha, S., Portela, C., Braz Junior, G., Almeida, J., Silva, A., ... & Rivero, L. (2022, June). Defining an A/B Testing Process for Usability and User Experience Evaluation Through the Analysis of the Results of a Literature Review. In International Conference on Human-Computer Interaction (pp. 204-213). Cham: Springer International Publishing.

[89] King, R., Churchill, E. F., & Tan, C. (2017). Designing with data: Improving the user experience with A/B testing. " O'Reilly Media, Inc.".

[90] Rivero, L., & Conte, T. (2017, October). A systematic mapping study on research contributions on UX evaluation technologies. In Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems (pp. 1-10).
[91] Hermawati, S., & Lawson, G. (2016). Establishing usability heuristics for heuristics evaluation in a specific domain: Is there a consensus?. Applied ergonomics, 56, 34-51.
[92] Nielsen, J., Clemmensen, T., & Yssing, C. (2002, October). Getting access to what goes on in people's heads? Reflections on the think-aloud technique. In Proceedings of the second Nordic conference on Human-computer interaction (pp. 101-110).
[93] Fan, M., Shi, S., & Truong, K. N. (2020). Practices and Challenges of Using Think-Aloud Protocols in Industry: An International Survey. Journal of Usability Studies, 15(2).
[94] Eriksson, Lina & Hermansson, Simon & Eriksson, Simon. (2010). Pinch analysis of Billerud Karlsborg, a partly integrated pulp and paper mill.
[95] Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial intelligence, 267, 1-38.
[96] Grice, H. P. (1975). Logic and conversation. In Speech acts (pp. 41-58). Brill.
[97] Frederking, R. (1996). Grice's maxims: do the right thing. Frederking, RE.
[98] Sanders, E. B. N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. Co-design, 4(1), 5-18.
[99] Page, G. G., Wise, R. M., Lindenfeld, L., Moug, P., Hodgson, A., Wyborn, C., & Fazey, I. (2016). Co-designing transformation research: lessons learned from research on deliberate practices for transformation. Current Opinion in Environmental Sustainability, 20, 86-92.
[100] VERBI Software. (2021). MAXQDA 2020 [computer software]. Berlin, Germany: VERBI Software. Available from maxqda.com.
[101] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.
[102] Yuasa, T., Harada, F., & Shimakawa, H. (2022, December). Proposal to Improve Exercise Using the Fogg Behavior Model. In 2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) (pp. 1-4). IEEE.
[103] Fogg, B. J. (2019). Fogg behavior model. Behav. Des. Lab., Stanford Univ., Stanford, CA, USA, Tech. Rep.
[104] BJ Fogg. Retrieved March 25, 2024, from https://behaviormodel.org/
[105] Harry Cloke. Retrieved March 25, 2024, from https://www.growthengineering.co.uk/bj-foggs-behavior-model/
[106] Ljungblad, S., Man, Y., Baytaş, M. A., Gamboa, M., Obaid, M., & Fjeld, M. (2021, May). What matters in professional drone pilots' practice? An interview study to understand the complexity of their work and inform human-drone interaction research. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-16).
[107] Crouch, M., & McKenzie, H. (2006). The logic of small samples in interview-based qualitative research. Social science information, 45(4), 483-499.
[108] Prpa, M., Fdili-Alaoui, S., Schiphorst, T., & Pasquier, P. (2020, April). Articulating experience: Reflections from experts applying micro-phenomenology to

design research in HCI. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-14).

[109] Speicher, M., Hall, B. D., & Nebeling, M. (2019, May). What is mixed reality?. In Proceedings of the 2019 CHI conference on human factors in computing systems (pp. 1-15).

[110] Dalton, J., & Dalton, J. (2019). Dot voting. Great Big Agile: An OS for Agile Leaders, 165-166.

[111] Spinuzzi, C. (2005). The methodology of participatory design. Technical communication, 52(2), 163-174.

[112] Plain, C. (2007). Build an affinity for KJ method. Quality Progress, 40(3), 88.

[113] Krug, S. (2014). Don't make me think, Revisited. A Common Sense Approach to Web and Mobile Usability.

[114] Hanington, B., & Martin, B. (2017). The pocket universal methods of design: 100 ways to research complex problems, develop innovative ideas and design effective solutions. Rockport.

[115] Noushad, B., Van Gerven, P. W., & de Bruin, A. B. (2023). Twelve tips for applying the think-aloud method to capture cognitive processes. Medical Teacher, 1-6.

[116] Stumpf, S., Rajaram, V., Li, L., Burnett, M., Dietterich, T., Sullivan, E., ... & Herlocker, J. (2007, January). Toward harnessing user feedback for machine learning. In Proceedings of the 12th international conference on Intelligent user interfaces (pp. 82-91).

[117] Billsus, D., Hilbert, D. M., & Maynes-Aminzade, D. (2005, January). Improving proactive information systems. In Proceedings of the 10th international Conference on intelligent User interfaces (pp. 159-166).

[118] Laredo, D., Qin, Y., Schütze, O., & Sun, J. Q. (2019). Automatic model selection for neural networks. arXiv preprint arXiv:1905.06010.

[119] Pu, P., & Chen, L. (2006, January). Trust building with explanation interfaces. In Proceedings of the 11th international conference on Intelligent user interfaces (pp. 93-100).

[120] Carenini, G., & Moore, J. (1998). Multimedia explanations in IDEA decision support system. In Working Notes of the AAAI Spring Symposium on Interactive and Mixed-Initiative Decision Theoretic Systems (pp. 16-22).

[121] Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019, May). Guidelines for human-AI interaction. In Proceedings of the 2019 chi conference on human factors in computing systems (pp. 1-13).

[122] McDougall, R. J. (2019). Computer knows best? The need for value-flexibility in medical AI. Journal of medical ethics, 45(3), 156-160.

[123] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., ... & Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. Stroke and vascular neurology, 2(4).

[124] Ferrario, A., & Loi, M. (2022, June). How explainability contributes to trust in AI. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 1457-1466).

[125] Massaro, A. (2022). Advanced control systems in industry 5.0 enabling process mining. Sensors, 22(22), 8677.

[126] Molenaar, N. J., & Smit, J. H. (1996). Asking and answering yes/no-questions in survey interviews: A conversational approach. Quality and Quantity, 30(2), 115-136.

[127] Pettersson, I., Lachner, F., Frison, A. K., Riener, A., & Butz, A. (2018, April). A Bermuda triangle? A Review of method application and triangulation in user experience evaluation. In Proceedings of the 2018 CHI conference on human factors in computing systems (pp. 1-16).

A

Appendix 1

A.1 Operator Interview Questions

Part 1 - Initial Questions

1. What does a normal workday look like for you?
   (a) How do you start your day?
   (b) When do you take breaks?
2. How does the bleaching/delignification process work, shortly summarized?
   (a) What variables in the process do you most often consider as an operator?
   (b) What are your goals in this process?
3. What computer system do you use in your work?
   (a) What is demanding in that system?
   (b) How do you overcome these demands?
   (c) What is the most relevant information from the system for your work?
4. How do you recover the system from problematic scenarios?
5. Do you have any recommendation systems currently?
   (a) What is good/bad with that recommendation system?
   (b) When you get a recommendation, how do you usually react? Do any questions arise?

Part 2 - Motivation

1. What are the easiest parts of your work?
   (a) Describe a situation where work felt easy.
2. What are the most tedious parts of your work?
   (a) Describe a situation where work felt tedious.
3. Which functionalities in the computer system are used the most?
   (a) Why?
4. Which functionalities in your system are rarely used even though they should be used?
   (a) Why?
5. Could you describe a situation where you felt like you improved in your work?
   (a) How/Why?
6. Could you describe a situation where you felt like work you did was especially meaningful?
   (a) How/Why?
7. Could you describe a situation where you were working and lost track of time?

   (a) How/Why?
8. What causes you to lose motivation in your work?
   (a) Describe a situation where you lost motivation in work.

Part 3 - Feedback

1. If you were to give feedback to a colleague on their/your own work, how would you do it?
   (a) Imagine that you and your colleague are sitting together in the control room. Your colleague suggests an action that you do not agree with, how would you handle that situation?
2. What response do you expect from the colleague?
3. As colleagues, when do you usually give feedback to each other?
4. When during your work shifts do you usually give feedback to each other?
5. How much time do you usually spend doing so?
6. Is there anything that encourages you to give feedback to your colleagues?
7. Are there any challenges in giving feedback?
   (a) Describe a situation where giving feedback was challenging.
8. Imagine you are giving feedback to a digital system instead of a colleague, how would you want to do that?
   (a) How do you expect the system to respond?
      i. Is there anything you need in return from the system when giving feedback?
   (b) How much time would you like to spend on giving feedback to the system?

Part 4 - AI Experience

1. Do you have any experience of using AI within your workplace?
   (a) How did you use the AI system, in what context?
   (b) What did you think of it?
   (c) Was it useful, why/why not?
   (d) Was there anything missing in the AI system?
   (e) Was it easy or difficult to understand?
   (f) Did you notice any issues?

Part 5 - Final Questions

1. Given the entire conversation we've just had, do you have any other questions or reflections?
2. The purpose of this work is to develop a new AI system that assists you in your work. As part of that process, we want to involve you operators in the actual design process. Would you be interested in participating in a design exercise sometime in the future?

B

Appendix 2

B.1 First Participatory Design Workshop Protocol

Aim

Generate design ideas for eXplainable AI (XAI) feedback interfaces that support motivation.

Introduction and Background

We are conducting this design workshop as part of our master's thesis, where the goal is to design an interface for giving feedback to an XAI system in an industrial production process. In this case, the production process is slow, and implemented changes takes a long time to be realized. To address this, ABB wants to implement an AI that produces a prediction for the process outcome. This will help the operators make decisions. However, the AI needs to be incrementally re-trained on-site, and should learn from the operator's expertise. The operators, who work in a large control room with desktop computers, should therefore be able to give feedback to the model, with the purpose of aligning the AI's predictions with their own. The feedback they provide gets interpreted and implemented into the model by a data scientist at a later point. Our goal is to ensure that the operators are motivated in giving feedback. How the operators can provide feedback, and how the AI responds to it, are important in this regard.

Insights into motivation from previous studies:

1. Operators get motivated by challenges and varied tasks
2. Motivation to give feedback depends on brief and easy to do interactions
3. Motivation to give feedback depends on how the system responds to feedback
4. Motivation to give feedback depends on AI usefulness
5. Motivation to give feedback depends on trust in the AI

Opportunities

1. The operators can choose the most accurate prediction from several different AI models.

2. What if the operators could combine AI models.
3. What if the operators could compete against each other to create the best combination.

Challenges

1. If the operators don't agree with a prediction, how do they communicate that?
2. If the operators combine models, how will they do it?
3. If the operators combine models, how can they explain their choices?

Procedure

1. Brainwriting 6-3-5 (30 min)
   (a) Each member writes/draws three ideas on one paper (5 min)
   (b) The paper gets passed to the next person who expands on those three ideas (3 min)
   (c) This gets repeated until everyone has contributed to every idea (15 min)
   (d) Everyone gets to briefly present their original ideas with added comments (10 min)
2. Rapid Paper Prototypes (15-20 min)
   (a) The group is divided into pairs.
   (b) Each pair creates a lo-fi paper prototype that can be tested (prepare tasks).
3. Testing (5-10 min)
   (a) One member from each pair gets to act as an operator and test another pair's interface using think-aloud.
4. Discussion of Pros and Cons (20-25 min)
   (a) Everyone writes down three pros and three cons for the designs on post-it notes (5 min)
   (b) Everyone gets a turn to explain their post-it notes (5 min)
   (c) Open discussion of the designs, brainstorms, and clustering of the post-it notes (10-15 min)
5. Dot voting (5 min)
   (a) Everyone gets three votes each and places them on their favorite brainwriting result and/or prototype (5 min)

C
Appendix 3

C.1 Evaluation Protocol

Introduction

Thank you for participating! Welcome to this session. We are students from Chalmers University of Technology and are conducting this as part of our master's thesis in collaboration with ABB. In this digital evaluation, which is estimated to take no more than an hour, you will get to test two different interfaces of a dashboard that uses an AI system to provide predictions for the kappa value (or brightness percentage) in the paper manufacturing processes.
First, you will read an introduction about each part of the dashboard so that you know what features are available. Feel free to ask questions if anything is unclear. After that, we will present you with several scenarios consisting of different tasks within the dashboard. For each scenario, we will provide instructions on how to perform the tasks.
This evaluation will be conducted using a technique called "Think-Aloud," which means you should verbalize your thoughts while performing the tasks. We are interested in your opinions and reflections during the session, so please share your impressions, expectations, and any aspects that you think work particularly well or poorly. Your feedback is crucial to us.
The session will also be recorded so that we can analyze the interface afterward and understand what has worked and what has not worked. Does this sound okay to you?

Version A

Scenario 1

1. In this scenario, you believe that the predicted Kappa value is too high. You want to communicate this to the computer. How would you proceed?
2. Now you want to create a comment. In your comment, you want to explain that the prediction relies too much on two sensors. How would you do this?
3. You also want to drag these two sensors here. How would you do that?
4. Now you can click on "Done".
5. What are your impressions regarding the confirmation dialog?
6. In this case, you cannot see your previous comments. What do you think about this?

Scenario 2

1. In this scenario, you know that it will rain very soon, and damp wood will enter the cooking process. You are aware that the damp wood will affect the Kappa value. However, the computer does not have access to this information, and you believe that the predicted Kappa value is too high. How would you proceed to incorporate this information into the prediction?
   (a) These bubbles can be seen as "obstacles" in the graph, and the prediction will attempt to avoid them. What are your thoughts on this?
2. You also know that two sensors will react specifically to the damp wood, and you want to include them in the "Moisture" section. By clicking on the icon in the bubble, a small window appears. How would you add the sensors there?
   (a) Close the window.
3. Now you want to inform the computer that this damp period will last for approximately 2 hours. You can achieve this by "stretching out" the bubble and making it elongated. What do you think about this approach?
4. You suspect that due to the damp wood, an H-factor adjustment will be necessary shortly after the rain begins. This adjustment is made to keep the kappa value at a slightly higher level. How would you incorporate this into the graph?
5. After making your adjustments, you want to inform the computer that the current prediction looks as expected. How would you do that?
6. Now that you are satisfied with your adjustments, you can click on "Done" in the corner.
7. The screen currently displays how it would look several hours into the future, after the entire rain scenario has played out. It turns out that the adjustments you made improved the prediction to better match reality. Do you feel that this is communicated effectively?
8. If you click the "NEXT" button, a new variant appears. Here, it shows that your adjustments did not align with reality. Do you think this is communicated well? Does it impact your motivation to provide additional feedback in the future?
9. If you wanted to remove one of the bubbles, how would you do it?

Version B

Scenario 1

1. Under the prediction graph, you see a tool for selecting different parts of the prediction graph. Ideally, you should be able to freely drag it and thus highlight various sections. Click anywhere on it and share your impressions.
2. Under this navigation tool, you see some blue boxes with comment icons. Click on the right one of the two and share your impressions.
3. In this scenario, you believe that the predicted kappa value is too high. You want to communicate this to the computer by creating a new comment. How would you do that?

4. You also want to write and explain that the prediction relies too much on two sensors. How would you do that?
5. You also want to drag these two sensors here. How would you accomplish that?
6. Now you can click on “Done".
7. If you wanted to revisit your comment and either view it or make changes, how would you do that?
8. If you wanted to delete the entire comment, how would you do that?
9. What are your thoughts on old comments being displayed like this?

Scenario 2

1. In this scenario, we have added an additional feature in the comment field that allows you to adjust the prediction yourself. So, you can start by creating a new “comment” and share your impressions about this new feature.
2. If you click “On”, you can interact with the prediction graph itself. You want to make a simple adjustment. How would you do that?
3. You feel satisfied with your adjustment and can click on “Done”.
4. If you wanted to revisit the comment and either view it or make changes, how would you do that?
5. If you wanted to delete your comment entirely, how would you do that?

Scenario 3

1. In this scenario, you know that it will rain very soon, and damp wood will enter the cooking process. You are aware that the damp wood will affect the kappa value. However, the computer does not have access to this information, and you believe that the predicted kappa value is too high. How would you proceed to incorporate this information into the prediction?
2. You also know that this damp period will last for approximately 3.5 hours. How would you communicate this to the computer?
3. You are also aware that a sensor will react specifically to the damp wood, and you want to include it in the “Moisture” section. How would you do that?
4. The value you currently see on the sensor is the current value. However, based on previous trends, you know that its value will be slightly lower when incoming wood is damp. How would you input this new value?
5. After entering the new value, the prediction was updated with this new information. What are your thoughts on this?
6. You also want to add another sensor and input a new value there. How would you do that?
7. Through the updated prediction, you notice that the predicted kappa value is too low. You see that an H-factor adjustment is needed to keep the kappa value stable. How would you incorporate this adjustment?
   (a) If you wanted to input an adjusted value for the H-factor, how would you do that?
8. Now you want to minimize the moisture box to take up less space on the screen. How would you do that?

9. Before clicking the NEXT button, take a look at the eye symbol. What do you think the eye symbol does? Experiment with it and share your impressions.
10. Now you can click the NEXT button. Several hours have passed. The screen now shows how it would look after the entire rain scenario has played out. It turns out that the adjustments you made improved the prediction to better match reality. Do you feel that this is effectively communicated?
11. If you click the NEXT button again, a new variant appears. Here, it becomes evident that your adjustments did not align with reality. Do you think this is communicated well? Does it impact your motivation to provide additional feedback in the future?

Scenario 4

1. In this scenario, you want to compare two models to see how their predictions differ. Model 1 has already been selected, and you want to add Model 2. How would you go about doing this?
2. What are your thoughts on viewing multiple graphs in this way?
3. How do you think you would want to interact with these two graphs? Are the features you previously went through still useful?

A/B Question

1. Now you have seen two versions of this dashboard, both with different ways to interact with the prediction. Which version do you prefer?

Summative Interview Questions

1. Is there any feature that stands out as particularly useful? Anything that feels redundant?
2. Regarding the ability to drag things into the graph and explore future scenarios, do you think it is a good feature?
3. Do you think it contributes to curiosity when interacting with the system?
4. Does this feature increase your understanding of how the AI model generates its predictions?
5. Would this feature encourage you to use the system more?
6. Do you think this feature could increase your trust in the system?
7. As you may have noticed, a green bubble has occasionally appeared with a "+1" or similar. Every time you provide input to the system, you accumulate points. If you click on the trophy icon in the navigation menu, you'll access a 'Scoreboard.' Click on it and share your impressions.
8. What do you think about if all operators collectively work towards milestones, and when these are achieved, you get treated to coffee?
9. If you click on the trophy icon again, you'll see another version where you work towards medals instead. Which version do you prefer?
10. What do you think about the color choices?
11. Do you believe this scoreboard is sufficient to explain how your feedback contributes to improved models and predictions? Is there anything missing?