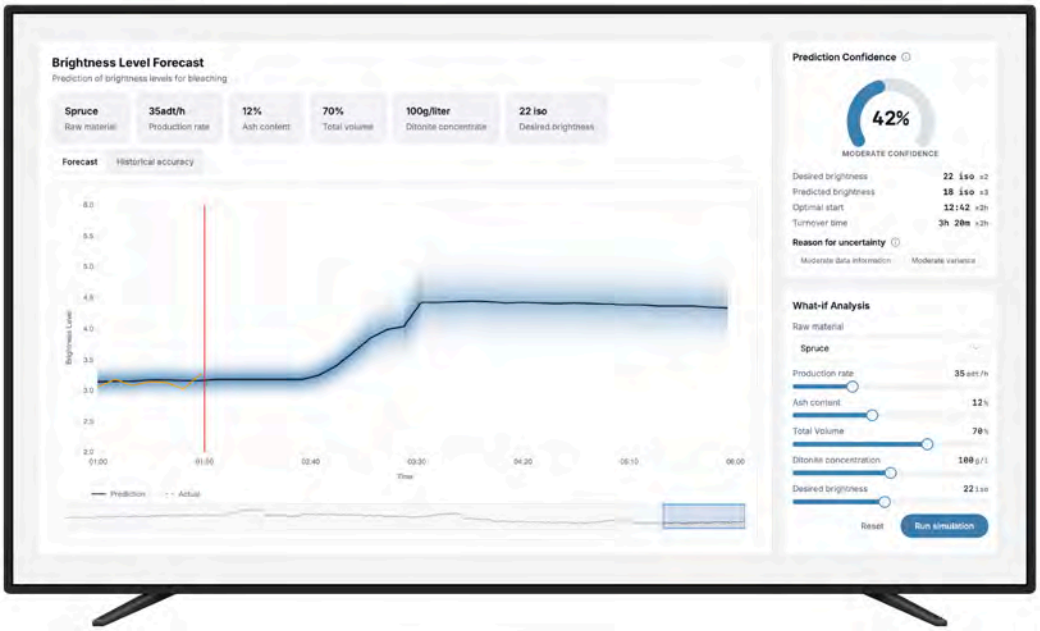
CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG



Designing for Uncertainty in AI: Supporting Decision-Making in the Process Industry

A User-Centered Approach to Communicating AI Uncertainty for Operators in the Process Industry

Master's thesis in Computer Science and Engineering

LISA LJUNGBERG & MAJA KJELLBERG

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2026

Master's thesis 2026

Designing for Uncertainty in AI: Supporting Decision-Making in the Process Industry

A User-Centered Approach to Communicating AI Uncertainty for Operators in the Process Industry

LISA LJUNGBERG & MAJA KJELLBERG

UNIVERSITY OF
GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2026

Designing for Uncertainty in AI: Supporting Decision-Making in the Process Industry
A User-Centered Approach to Communicating AI Uncertainty for Operators in the Process Industry
LISA LJUNGBERG AND MAJA KJELLBERG

Supervisor: Ilaria Torre, Department of Computer Science and Engineering
Advisor: Mowei Yang, ABB
Examiner: Palle Dahlstedt, Department of Computer Science and Engineering

Master's Thesis 2026
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Dashboard for paper mill bleaching operators, translating design guidelines into an interactive interface prototype.

Typeset in LaTeX
Gothenburg, Sweden 2026

Designing for Uncertainty in AI: Supporting Decision-Making in the Process Industry
A User-Centered Approach to Communicating AI Uncertainty for Operators in the Process Industry
LISA LJUNGBERG & MAJA KJELLBERG
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

This thesis investigates how uncertainty in artificial intelligence (AI) systems can be effectively communicated to support decision-making in the process industry. As machine learning-based decision support systems are increasingly integrated into industrial environments, operators are required not only to interpret system outputs but also to assess their reliability. However, current approaches to uncertainty communication are often insufficiently intuitive, limiting trust and appropriate reliance on AI systems.

Adopting a human-centered design approach, this study explores how industrial operators perceive and reason about uncertainty in their everyday work, and how different forms of uncertainty communication influence trust, decision-making, and system use. The research is conducted within the context of the pulp and paper industry, with a particular focus on the bleaching process, a complex and safety-critical operation.

The study follows a Double Diamond design process, combining methods such as a literature review, semi-structured interviews, and observations. Based on these insights, a set of design guidelines for uncertainty communication is developed, validated through a design workshop, and implemented in a high-fidelity user interface prototype. The guidelines are then evaluated through the prototype with domain experts using think-aloud protocols, interviews, and acceptance measures.

The findings show that operators rely heavily on experience-based and rule-based reasoning when handling uncertainty, and that transparency, contextual explanations, and historical performance data are essential for building trust in AI systems. Furthermore, effective uncertainty communication encourages more reflective decision-making and supports appropriate reliance on AI recommendations.

This thesis contributes with empirically grounded design guidelines and a validated prototype that demonstrate how uncertainty can be communicated in a way that aligns with operators' cognitive processes and work practices.

Keywords: Artificial Intelligence (AI), Uncertainty communication, Explainable AI, Human-AI interaction, Process industry, User-centered design, Design guidelines

Acknowledgements

We would like to express our gratitude for the opportunity to conduct our masters thesis at ABB Corporate Research. We are especially thankful to our industry advisor, Mowei Yang, for her continuous support, encouragement, and invaluable guidance throughout the entire project. Her insights into the design process, as well as her thoughtful recommendations on how to approach and conduct user studies, have been important to our work and learning.

We would also like to extend our thanks to Holmen Paper and Board Braviken Paper Mill and the operators who generously dedicated their time to participate in our interviews and evaluation sessions. Their openness, expertise, and willingness to share their experiences provided invaluable insights that shaped our design process and formed our designed guidelines.

Finally, we would like to sincerely thank our academic supervisor, Ilaria Torre, for her continuous guidance, patience, and support throughout this thesis. Her constructive feedback and academic expertise have been essential in improving the quality of our work and guiding us through the research process.

Lisa Ljungberg & Maja Kjellberg, Gothenburg, 2026-06-04

Contents

List of Figures

List of Tables

1

Introduction

Recently, information technologies such as artificial intelligence (AI) have taken on an increasingly larger role across all stages of process industries [1]. In particular, machine learning-based decision support systems (ML-DSS) are now deployed in high-stakes operational environments, where they aim to reduce cognitive load and assist complex decision-making. Yet AI-generated outputs inherently involve uncertainty. Operators must therefore interpret not only the systems recommendations but also the level of confidence associated with those predictions [2].

Although ML-DSS offer substantial potential benefits, their limited transparency can mislead non-expert users. When uncertainty is hidden or poorly communicated, users may develop overconfidence and over-rely on recommendations that are biased or inaccurate [3]. While uncertainty communication is increasingly recognized as central to fostering trust in AI systems, there remains limited knowledge about how industrial operators actually reason about and respond to uncertain AI predictions in real-world operational contexts [2].

Humans are accustomed to making decisions under uncertainty, but this familiarity can sometimes foster over-confidence particularly when evaluating AI-generated outputs [4]. In industrial environments, where decisions are often time-critical and safety-sensitive, misinterpreting uncertain AI outputs may have serious consequences. A deeper understanding of how uncertainty should be communicated and designed in human-AI interfaces is therefore essential.

This thesis builds on the completed EXPLAIN project (EXPLanatory Interactive Artificial Intelligence for INdustry), an international collaboration between industry partners (ABB, Boliden, Södra, LEAG, Prodrive Technologies, Mek Europe BV, and others) and universities (Umeå University, Linköping University, University of Hildesheim, Technische Universität Darmstadt, and Delft University of Technology) [5]. The EXPLAIN project focused on developing human-centered, explainable AI (XAI) for process industries and established a framework that prioritizes user needs over purely technical transparency. It provides practical strategies and real-world examples to support AI developers, engineers, and decision-makers in integrating explainable AI into industrial contexts. Within this framework, communicating prediction uncertainty is considered a core principle, enabling users to assess AI reliability and build appropriate trust [5].

Despite this recognition, empirical research on how uncertainty should be effectively

communicated in process industries remains limited. This thesis addresses that gap by examining how uncertainty in AI systems can be communicated to increase safe and informed decision making within process industries, with a specific focus on paper mills.

The pulp and paper industry is a well-established and economically significant sector in Sweden, with 38 active paper mills nationwide. As one of the worlds leading exporters of pulp and paper [6], [7], [8], Sweden relies heavily on highly automated production environments. Nevertheless, human operators continue to play a crucial role in supervising both automation and process performance [9]. Designing technical systems for such environments therefore requires careful consideration of the interaction between operators and automation, particularly in supervisory control settings [10]. Studies show that operators tend to perceive automation as effective and supportive during stable operations. In contrast, during disturbances, trust in automation and its perceived usefulness often decline [10].

The relationship between operators and automation in paper mills is thus multifaceted. Automation can enhance efficiency and reduce workload, yet it may also generate frustration in adverse situations, especially when operators struggle to determine whether the system is functioning correctly [10]. At the same time, paper mills manage large volumes of time-series data, increasing the importance of accurate and reliable AI outputs [5].

Although uncertainty visualizations has advanced in recent years, most studies have focused on general tasks or controlled experimental settings. Consequently, there is limited understanding of how industrial operators experience and manage uncertainty in their everyday work. As AI-based prediction systems are increasingly considered for implementation in process industries, understanding operators existing uncertainty reasoning becomes critical. The EXPLAIN project addressed this need by developing uncertainty visualizations for industrial predictions; nevertheless, operators reported difficulties interpreting them [5]. Building on these findings, this thesis investigates how operators currently interpret and act upon uncertainty without AI support. Furthermore, it explores how interaction design can inform the development of design principles for effectively visualizing and communicating uncertainty in future AI prediction systems.

1.1 Research Questions

1. **RQ1** How do industrial operators interpret and reason about uncertainty in their current work practice?
2. **RQ2** How does the communication of uncertainty influence operators trust, reliance, and decision-making under uncertainty?
3. **RQ3** What design guidelines can support the operator to trust and understand the level of uncertainty in AI outputs?

1.2 Goals and Deliverables

This work aims to bridge the gap between Human-Centered Explainable AI principles and the practical visualization of uncertainty in industrial settings. Building on the EXPLAIN project, we seek to develop design guidelines and interface prototypes that effectively communicate AI prediction uncertainty to process industry operators.

The expected deliverables for this project are;

- **Design guidelines** for communicating uncertainty in AI for paper mill operators.
- **User interface prototype** implementing the design guidelines to demonstrate effective uncertainty communication in operator interfaces.

1.3 Stakeholders

This thesis is conducted within the Interaction Design and Technologies master's program at Chalmers University of Technology, in collaboration with ABB Corporate Research. The academic supervisor is Ilaria Torre, with Mowei Yang serving as the industrial supervisor from ABB. The work focuses on the process industry, specifically the pulp and paper industry, where control room operators represent key stakeholders whose needs and practices inform the design process.

2

Background

This chapter establishes the context for the thesis. It begins by presenting the industrial context of the work, followed by an examination of Artificial Intelligence (AI) integration in the process industry. The chapter then discusses the role of operators in AI-supported systems. Finally, it outlines key principles of uncertainty communication and reviews related research on uncertainty in AI.

2.1 The Industrial Context

This thesis focuses on uncertainty communication within the process industry, specifically in the pulp and paper domain. The following section introduces the EXPLAIN project, which this thesis builds upon and presents the use case for this work: the paper bleaching process.

2.1.1 EXPLAIN Project

This thesis builds on the EXPLAIN project (EXPLanatory Interactive Artificial Intelligence for INdustry), an EU funded project that investigates how to design artificial intelligence through a human-centered approach for the process industry [5].

A key challenge addressed by the EXPLAIN project was the so-called 'black box problem' in industrial AI. AI is increasingly introduced into the process industry, which includes pulp and paper, chemicals, oil and gas, cement, metals, food, and beverages. In these types of industries operators often find the systems difficult to understand and therefore difficult to trust. This creates a barrier to both adoption and effective use of AI in environments where decisions are complex and risky [5].

EXPLAIN therefore aimed to develop user-friendly, transparent and supportive AI systems that enhance operators work rather than replace their expertise. The goal was to create AI solutions that provide understandable explanations, increase the users understanding of the model's decisions, and facilitate close collaboration between humans and AI systems [5].

The project started from the principles of Explainable AI (XAI), but developed them further into a human-centered explainable AI framework (HCXAI). This framework emphasizes that explanations should not just be technical visualizations of the mod-

els internal logic, they should be social, that is, designed to meet human needs, language, goals, and work context. Explanations should help the user understand why an AI decision was made, how it relates to their mental models, and what can be done about it [5].

By working closely with operators, engineers, and other domain experts, EXPLAIN demonstrated that explanations need to support operators workflows, their risk assessments, and their goals, rather than stand alongside them as separate technical components. The focus was therefore on interactivity, contextualization, uncertainty communication, and that explanations should support real-time decision-making [5].

The project clearly demonstrated that communicating uncertainty in AI predictions is crucial for enabling operators to make informed decisions. At the same time, EXPLAIN highlighted a continued need to further develop methods for communicating uncertainty in industrial AI systems, as the project's proposed visualization designs for uncertainty were often perceived as insufficiently intuitive for users [5].

2.1.2 Pulp Bleaching Process

The central use case of this report is the bleaching process within paper and pulp production. The description will be presented below.

At Holmen's paper mill, the entire production process is managed at one site, from wood chips to finished paper. The pulping stage uses a method called **Thermomechanical Pulping (TMP)**, where wood chips are mechanically processed into pulp. After production, the pulp is stored in silos before proceeding to **the bleaching process**.

The bleaching process is the central use case of this report. It consists of distinct phases that together transform the mechanically produced pulp into bleached pulp with a specified brightness level. The entire process is monitored and operated from a DCS-system (Distributed Control System). Drives, instrumentation and valves are integrated into the DCS-system. With help of the DCS system, the operators are able to start and stop the process equipment and set targets for controllers (flow, pressure, etc).

In the following sections, each phase of the pulp bleaching process will be explained in detail, additionally a simplified model of the bleaching process is included (Figure 2.1).

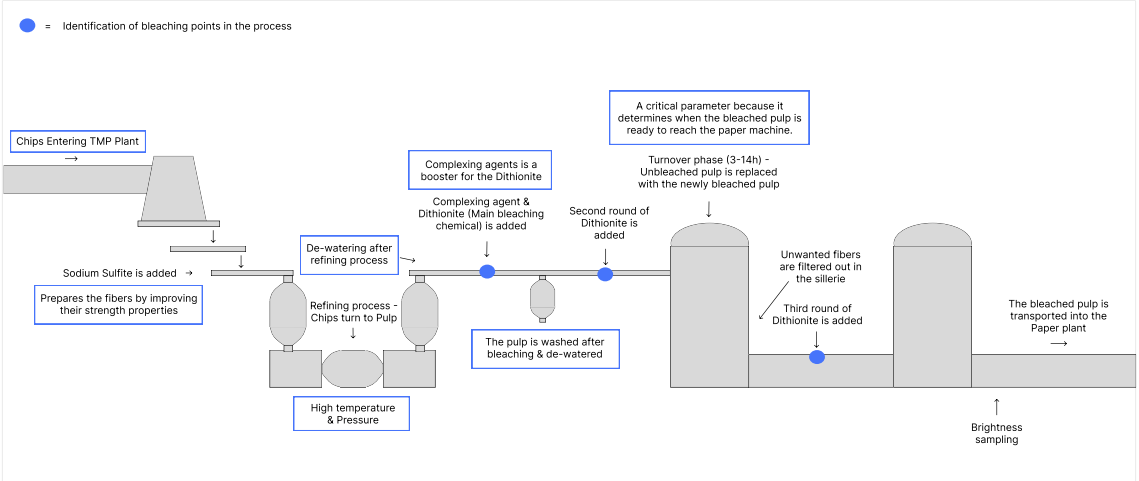


Figure 2.1: Overview of the Thermomechanical Pulping (TMP) process

2.1.2.1 Preparation for Bleaching

Before the pulp enters the actual bleaching process, it undergoes preparation. The wood chips are screened and stored in separate silos according to wood type in the chip handling system. As the chips are processed into pulp, a chemical pretreatment is initiated: sodium sulfite is added in the pressurized conveyors ahead of the refining. This chemical prepares the fibers by improving their strength properties and contributing to a minor brightness increase.

After chemical pretreatment, the chips enter the refining system. Under high pressure and temperature, they are refined into pulp, and the fibers are developed to achieve the desired mechanical and optical properties. This phase establishes the foundation for effective bleaching.

This pretreatment step bridges the TMP production and the bleaching process, ensuring that the pulp is optimally conditioned before entering the bleaching sequence.

2.1.2.2 Primary and Secondary Bleaching

Once the pulp has passed through the refining system, steam is removed from the pulp and the primary bleaching stage begins. The operator initiates the dosing of dithionite into the standpipes at a ratio of 5 kg per ton. A complexing agent is introduced concurrently to enhance dithionite stability, increase bleaching efficiency, and reduce associated odors

Dithionite is the key bleaching chemical and can increase brightness by up to 14 ISO units. Before the transition to higher brightness, an additional dithionite dose is added. This secondary dose provides an extra boost to ensure the pulp reaches the target brightness.

2.1.2.3 Bleaching Turnover Control

After the bleaching chemicals are added, the pulp stored in pulp towers. During this storage period, the old, unbleached pulp that was already in the towers is gradually replaced by the newly bleached pulp, known as 'turnover'. The turnover time is a critical parameter because it determines when the bleached pulp is ready to reach the paper machine. Currently, operators use a turnover table (i.e., recipe) based on the total volume percentage in the pulp towers, where turnover time varies from 3 hours at 40% volume to 14 hours at 190% volume.

Once the turnover is completed, the pulp brightness is evaluated to determine whether the bleaching process has produced the desired result. At this stage, the pulp may meet the target brightness, be overbleached, or be under-bleached. If brightness is too high, operators can compensate by adding dyes for reduction of brightness. However, if brightness is too low, the pulp cannot be sent to customers, potentially leading to production stoppages and financial losses.

After the bleaching process is complete, the pulp is fully prepared for conversion into paper.

2.1.2.4 Challenges of the Bleaching Process

In the turnover phase of the bleaching process, a significant challenge becomes apparent. To avoid the risk of insufficient bleaching, the mill operates with a deliberate safety margin: it is considered better to bleach too early rather than too late. Overbleaching can be corrected later with a small amount of dye at a relative low cost, whereas under-bleaching may lead to paper with too low brightness B-grade with severe economic consequences. This proactive approach ensures stable operations and reliable deliveries to customer.

However, even with clear guidelines and built-in safety margins that encourage early bleaching, this risk-averse strategy introduces its own drawback: chemicals may be used in excess. Although this ensures uninterrupted production, it can also result in unnecessary chemical consumption, increasing both operational costs and environmental impact. The central challenge, therefore, is to optimize the bleaching process so that the required brightness is achieved while avoiding avoidable chemical overuse. The goal is to maintain process reliability without compromising cost efficiency or sustainability.

2.2 AI Integration in Process Industry: Opportunities and Challenges

Artificial Intelligence has had a strong impact and is being integrated into more and more industries to increase efficiency [1], [5]. In the process industry, the benefits are particularly great because AI can optimize production processes, reduce deviations and minimize waste [5]. As automation has become more advanced, significantly fewer operators are required to run a process. However, despite this increased level

of automation, several factors still prevent the process industry from becoming fully autonomous, and competent personnel remain a cornerstone of successful factory operations [9].

Even though AI provides clear advantages, its integration also brings notable challenges. One of the most critical is the black box problem, where AI systems generate results or decisions without offering transparent insight into how these conclusions were reached [5]. In safety-critical and heavily regulated environments such as the process industry, this lack of transparency becomes particularly sensitive. It can reduce operators trust in the system, limit the usability of AI tools, and make integration into existing workflows more difficult. To enable broader and more effective use of AI, there is therefore a need for methods that increase explainability and help operators and decision-makers understand model behavior [5].

Beyond the issue of transparency, AI systems often produce predictions that contain some degree of uncertainty. This means that operators must not only interpret what the AI recommends, but also understand how confident the model is in its predictions [2]. Handling this uncertainty becomes an additional cognitive task and further emphasizes the importance of human expertise. Another key aspect in the process industry is the type of data being analyzed. Time series data is the most central data type, forming the basis for monitoring, controlling, and optimizing industrial processes. Such data is characterized by very large data volumes, strong temporal dependencies, multivariate relationships between many measurement variables, the need for real-time anomaly detection, and the possibility to extract historical insights to support future decision-making [5]. These characteristics make time series data quite complex, reinforcing the need for AI systems that are not only powerful but also transparent and supportive of the human operator.

2.3 The Human Operator

Human operators play a central role in industries such as pulp and paper manufacturing. In modern plants, their primary responsibility is to supervise automated systems and processes [9]. In addition, process operations often involve multiple operators working simultaneously, requiring ongoing coordination, maintenance activities, and effective communication within teams and across shifts [9]. However, operators in control rooms tend to prioritize performance and operational efficiency [11]. In addition, some operators are reluctant to interact directly with the process, as they are aware that mistakes can lead to significant financial losses or compromise process safety [9].

The control room operator oversees plant operations by analyzing both forecasts and real-time data. Their role involves maintaining process stability, ensuring product quality exceeds target levels, and proactively making adjustments to optimize performance. They also address challenges such as declining data quality and fault detection [5].

Engineers can often overlook the role of human operators. Gaining insight into human behavior and capabilities is essential for creating more effective human - system

interfaces. Furthermore, operators are highly experienced users with specific needs, which makes their involvement in interface design essential [9]. This underscores the importance of considering the operator throughout the entire design process.

2.3.1 Human Operator in Control Rooms

In control rooms, operators typically interact with many different interfaces [9]. And the operation of control rooms necessitates high demands from both human operators and technological infrastructure to ensure optimal functionality. As control room operators across various industries assume increasingly complex responsibilities in the future, these operational demands are expected to become more challenging [11].

2.3.2 Cognitive Considerations

To understand the operators further we have highlighted some cognitive considerations to have in mind.

Rasmussens Skill-Rule-Knowledge (SRK) framework provides a useful lens for describing how human behavior varies depending on familiarity, task demands, and situational complexity [12]. Atkinson and Hollender highlight that this framework is particularly relevant in modern process industries, where automation has shifted the nature of operators tasks toward higher-level cognitive reasoning [9].

1. **Skill-based behavior** refers to actions that are executed automatically with little or no conscious attention. In industrial plants, many tasks that once required skill-based human performance have been automated, meaning operators are less frequently engaged at this level. As a result, most operational tasks today demand behavior at higher cognitive levels [9].
2. **Rule-based behavior** is applied in situations that are familiar but not fully automatic. Operators select actions by applying set rules, these rules may be learned through experience or provided from previous operators [9].
3. **Knowledge-based behavior** becomes necessary when operators encounter new or unexpected situations. In such cases, operators must rely on an understanding of the principles and dynamics of the process to diagnose the situation and formulate appropriate goals. Because of this, the cognitive workload of the operator is higher here compared to the other levels [9].

Another key point to consider is that human short-term memory has a limited capacity, typically around 7+-2 items [13]. This implies that Human System Interaction (HSI) designs should avoid overloading operators by requiring them to remember too many process values at once [9].

2.4 Uncertainty Communication

Incorporating uncertainty communication is important for effective human-AI interaction. Prior research shows that meaningful transparency achieved through clear and informative explanations of AI reasoning can reduce perceived uncertainty, increase user trust, and improve willingness to rely on AI-supported decisions [14]. Furthermore, presenting uncertainty in machine learning (ML) predictions encourages users to pause and engage in more careful, analytical decision-making rather than relying on automated outputs uncritically [2]. To better understand why such communication is essential, it is necessary to consider the nature of uncertainty in AI systems themselves.

AI models produce uncertain predictions, due in part to limited or incomplete data, the models inherent limitations, and the complexity of the real-world phenomena they attempt to capture [2]. While humans are accustomed to making decisions under uncertainty, this can lead to overconfidence in their ability to interpret AI systems outputs, risking a false sense of security regarding the models predictions [4]. Results from the EXPLAIN project clearly demonstrate that communicating such uncertainty is crucial for operators to make well-informed decisions; at the same time, current methods for visualizing uncertainty in industrial AI systems are often perceived as insufficiently intuitive and therefore need further development [5]. Moreover, since many AI models present their predictions without any indication of their reliability, there is a real risk of misleading users. To enable more informed and responsible decision-making, prediction systems should therefore explicitly communicate the degree of uncertainty, thereby helping operators better assess the reliability of the models outputs [5]. It is worth noting that uncertainty communication is a broad concept, encompassing a range of modalities beyond the visual, however, this project focuses specifically on visualization of uncertainty.

The following sections offer an in depth discussion of uncertainty communication and visualization in various contexts, as well as a presentation of visualization techniques relevant to time series forecasting.

2.4.1 Challenges in Understanding Uncertainty

Research has shown that numeracy influences how people evaluate and respond to risk information [15]. Many people struggle to interpret numerical data and to grasp the uncertainty it conveys, highlighting the need for alternative approaches to communicating uncertainty to individuals with lower numeracy skills. Uncertainty communication techniques to address this issue could include explaining the level of uncertainty in words and presenting the uncertainty as a numerical range [16].

In addition to numeracy, cognitive biases also shape how uncertainty is understood. One prominent example is the framing effect: individuals tend to respond more favorably to information presented in a positive frame than to the same information presented in a negative frame. This preference can substantially influence how risks and uncertainties are perceived. To mitigate framing effects, it is important to present uncertainty using both positive and negative outcome descriptions, thereby

supporting a more balanced and informed interpretation [16].

People interpret uncertainty in different ways depending on various factors, including past experiences, personal biases, knowledge, and skill levels. Therefore, when designing for uncertainty, it is essential to clearly understand who the intended users are. Designers must consider who the users are, their expectations, needs, and constraints to create solutions that support informed decision-making and reduce confusion [16].

2.4.2 Uncertainty Visualization Techniques

Visualization of uncertainty should be context-sensitive and user-specific [17]. However, when looking beyond the process industries domain, AI uncertainty has been explored in other areas such as gaming and healthcare.

Reyes et al. explored how uncertainty in AI can be visualised in the context of gaming, predictions, and decision-making, and examined the impact of such visualisations. Their findings showed that visualising an AI systems uncertainty increased participants trust, particularly among those who were initially sceptical or negative toward the use of AI. Additionally, the authors provide recommendations for AI designers, emphasizing that clearly communicating AI uncertainty can encourage more informed and reflective decision-making. Furthermore, they recommend considering human factors such as visual perception when communicating AI prediction uncertainty [18].

In the medical AI domain, uncertainty has also been studied. Cao et al. investigated factors influencing human-AI collaboration in a skin cancer screening task. Their findings indicate that presenting calibrated model uncertainty alone is insufficient. However, presenting calibrated uncertainty in a frequency format - for example, showing how often the AI was correct, helps users adjust their reliance more effectively and reduces confirmation bias [19].

2.4.3 Uncertainty Visualization Techniques for Time Series Forecasting

To make time series models understandable to both experts and novices, a number of visualization techniques have been developed. Many of these models, especially deep neural networks, are difficult to interpret, and research on this has focused on two main avenues for visualization: exploring explanations for models that are already relatively easy to understand, such as decision trees, or creating new explanatory mechanisms that can provide insight into more complex "black box" models. Which method to use depends on the context, the users needs and workflow. Different domains and purposes require different types of explanations, which is why a user-centered approach is extra essential [17].

2.4.3.1 Recommendations for enhancing Uncertainty Communication in Time Series Predictions

Karagappa et al. studied how different types of time-series prediction visualizations affected user performance, focusing on public-facing epidemiology dashboards. Additionally, the authors note that these guidelines can be applied in other contexts [20]. The visualization types they evaluated included confidence bands, overlapping bands, blur, circular glyphs, and colored markers [20]. Their results showed that colored markers performed the worst, while confidence bands offered no significant advantage and did not effectively convey the qualitative aspects of a credible interval. They also observed a strong relationship between perceived clutter and the aesthetic appeal of the visualizations.

To better understand user needs, the authors conducted a follow-up study comparing the best performing uncertainty visualizations: overlapping bands, blur, and circular glyphs. Circular glyphs outperformed the others in terms of task performance. Regarding information needs, about one third of uncertainty estimation tasks revealed that participants lacked sufficient information to make accurate estimates. The correlation between perceived clutter and aesthetics was confirmed in this study as well [20].

Based on these findings, Karagappa et al. proposed guidelines for designing uncertainty visualizations, aimed at improving clarity, usability, and user comprehension [20].

1. **"Standardizing Uncertainty Terminology and Visualization Techniques"**. There is currently no consistent way to describe uncertainty in predictions - terms like Credible Interval, Prediction Interval, Simulation Percentile, and Uncertainty Interval are all used. The choice of term often depends on the modeling method, but this inconsistency can be confusing for both researchers and users of visualizations. When showing uncertainty in a visualization, its important to make it both clear and accurate. The way uncertainty is displayed, and the terms used to describe it, should correctly represent whether its a precise range, a probability interval, or a more complex distribution [20].

2. **"Meeting the Informational Needs of a Diverse Audience"**. To ensure the users can make an informed decision it is important to give the users more **statistical information** to help them link the likelyhood and uncertainty intervals. Additionally, more information about **model information** is needed such as parameters and historical accuracy. Showing previous forecasts alongside actual outcomes could help communicate historical accuracy. Additionally, features like tool-tips, hover-over explanations, annotations, or narrative descriptions can help with this [20].

3. **"Equalizing Effects of Numeracy in Uncertainty Estimation"**. Reducing the mental effort required to make multiple inferences is especially helpful for users with lower numeracy skills. To support comprehension, designers should avoid visual elements that don't convey relevant information, as removing unnecessary details helps users focus on what matters. Additionally, use consistent and easy-to-

understand formats, for example, simplify uncertainty ranges so they are clear and manageable [20].

4. **"Clutter Reduction and Aesthetic Design Influence on Uncertainty Estimation"**. When designing visualizations, it is important to clearly show uncertainty ranges and avoid overlapping them. Poor handling of these elements can lead to confusing axis adjustments and visual clutter, which reduces the overall effectiveness of the chart [20]. To make graphs easier to interpret, unnecessary elements like backgrounds, borders, gridlines, and extra axis lines should be removed, spacing between key elements should be increased, and legends can be replaced with direct labels [21].

Kay et al. examine uncertainty visualization in a public transportation mobile application, and they explain how users interpret, understand, and use predictive uncertainty in time-pressured and information-intensive situations. The study shows that users often misinterpret point forecasts as more accurate than they actually are, which can lead to incorrect decisions. They argue that a combination of point estimates and explicit uncertainty representations reduces the risk of false precision, and that users can gain a clearer and more realistic picture of the potential risk distribution [22].

Additionally, Padilla et al. emphasize that there is no single visualization technique that works for all uncertainty cases. Designers need to consider each design decision, as poor choices can add confusion and hinder the decision-making process [23].

In our study, we will investigate these guidelines further in the process industry, specifically examining how they influence expert operators understanding, reliance, and trust in the visualizations.

3

Theory

The following section presents the theoretical foundation for this work. The theories are organized into three main areas: Artificial Intelligence, Cognitive frameworks, and Design theory. Each area provides essential context and concepts that support the analysis and discussion in later chapters.

3.1 Artificial Intelligence (AI)

This section introduces artificial intelligence and machine learning concepts relevant to industrial decision support systems, with particular focus on predictive analytics.

Artificial intelligence is increasingly being integrated into process industries to enhance efficiency and minimize waste. AI systems in these contexts analyze large volumes of multivariate time series data and serve as decision support tools for operators [24]. A key technology enabling these capabilities is machine learning, which forms the foundation for predictive analytics in industrial settings.

3.1.1 Machine Learning

Machine learning (ML) is a subfield of artificial intelligence that focuses on developing statistical algorithms with the ability to learn from data and generalize to new, previously unseen information [25].

Machine learning models function as flexible statistical frameworks that identify patterns in data and adjust their internal parameters based on those patterns. The field encompasses a wide range of methods, such as k-nearest neighbors (k-NN), clustering algorithms, and advanced neural networks. These models are applied in various tasks -including classification, regression, prediction, optimization, and pattern recognition and play a central role in many modern technological systems [25].

Within process industries, ML models are particularly valuable for predictive analytics, which is the primary focus of this work.

3.1.1.1 Predictive Analytics ML

Predictive analytics AI leverages machine learning algorithms and statistical methodologies to analyze patterns within historical and real-time data to forecast future

events. Through training on extensive datasets, these algorithms identify temporal trends, uncover correlations, and generate predictive outputs that facilitate decision support for operators, enable process optimization, and provide capabilities for anomaly detection [26].

However, the reliability of predictive analytics depends critically on data quality and model maintenance. Predictive models require training on accurate and well-structured data; incorrect values or missing data can rapidly degrade prediction accuracy. Furthermore, these systems must be monitored and updated continuously, as their performance may deteriorate when conditions in the operational environment change [26]. This inherent uncertainty in ML predictions - stemming from data limitations, model constraints, and environmental variability - creates challenges for operators who must decide when to trust and act upon AI predictions.

3.2 Cognitive Frameworks

The following section covers cognitive theories relevant to this thesis. Understanding how operators construct mental models, process information, and make decisions under uncertainty is critical for formulating guidelines and designing AI interfaces that support effective human-AI collaboration in industrial settings.

3.2.1 Cognitive Load Theory

Cognitive Load Theory, introduced by John Sweller, explains how human working memory processes information and emphasizes its limited capacity [27].

There are different types of cognitive load. Intrinsic cognitive load refers to the inherent difficulty of a task, meaning that some activities naturally require more mental effort to understand due to their complexity. In contrast, germane cognitive load relates to the mental effort involved in processing, organizing, and integrating new information into long-term memory through the development of schemas [28].

In the process industry, operators are often required to process large amounts of information on a daily basis. Therefore, it is important to design interfaces in a way that avoids overwhelming users with unnecessary information.

3.2.2 Mental Models

In a complex world, humans construct internal representations of external reality, referred to as mental models. These models provide simplified frameworks for understanding how the world operates, supporting decision-making and problem-solving processes. Because mental models are shaped by individuals prior experiences, knowledge, and perceptions, they naturally vary from person to person. However, they seldom offer fully accurate representations of the external world, as they are constrained by contextual factors and the subjective nature of personal experience [29].

A crucial aspect of this work is to explore and understand operator's mental models, ensuring that proposed design solutions align with their existing cognitive frameworks.

3.2.3 Decision Making Processes

Decision making refers to a cognitive process including, perception, memory, and attention. It also involves the process of making a choice. Decisions are shaped by processes that happen both before the choice, such as evaluation and judgment, and after the choice, such as feedback and learning [30].

Decision making involves perceiving and evaluating a large amount of information, and several key factors influence this process. One important factor is **uncertainty**, particularly the extent to which outcomes are unknown. When potential consequences are uncertain and may be negative or costly, the decision is often experienced as risky. **Time** is another critical element in decision making. This includes both the moment at which a decision is made and the presence of time pressure. **Familiarity** and **expertise** also play a significant role. Individuals with extensive experience in a domain may quickly recognize the best option, whereas less experienced individuals often find the same decision more challenging [31].

3.2.3.1 Information Processing of Decision Making

When making a decision, the decision-maker looks for cues of information from the environment that can guide their choice. These cues often contain uncertainty. Selective attention plays a crucial role in deciding which cues to focus on and which to ignore. This process is influenced by past experiences stored in long-term memory and requires effort and attentional resources. Once cues are attended to, they are perceived as part of the decision-makers understanding of the situation. At this stage, the decision-maker forms hypotheses about the current and potential future state of the world. This evaluation draws on both external cues filtered by selective attention and prior knowledge from long-term memory. What distinguishes decision-making is that these situational assessments are often imperfect or inaccurate [31].

Additionally, many decisions are iterative: initial hypotheses can prompt the search for additional information to either confirm or challenge them, creating a feedback loop. Finally, the decision-maker selects a course of action. When the state of the world is uncertain, this choice involves considering the potential consequences of each option [31].

In our work, understanding human decision-making is essential. People are accustomed to making choices under uncertainty, which can sometimes lead to overconfidence in their decisions about interpreting AI predictions [4]. As a result, it is crucial to explicitly design for uncertainty in AI systems so that users can make more informed decisions.

3.3 Design Theory

This section presents the design theories and methodologies that guide the development of user interfaces for communicating AI uncertainty. It covers human-centered design principles, participatory approaches, and the design process framework used throughout this work to ensure that proposed solutions align with operator's needs and work practices.

3.3.1 Human-Centered Design

Human-centered design is a widely used approach within UX design that ensures proposed solutions genuinely address the needs and requirements of people, particularly the intended users. A key aspect of this approach is developing a deep understanding not only of the users themselves but also of the context in which they interact with a product or service [32]. The human-centered approach is grounded in four core principles, outlined below.

1. **People-centered:** The design process should focus on people and the contexts in which they live and work. Various methods can support this, including participatory design, which actively involves users throughout the design process. This ensures that proposed solutions align with users lived experiences and contextual realities[32].
2. **Understand and solve the right problem:** Effective design requires a thorough understanding of the problem space. Without a well-defined problem, solutions risk addressing symptoms rather than root causes, leading to recurring issues [32].
3. **Everything is a system:** Designers should recognize that every element is part of a larger, interconnected system. Understanding these relationships helps ensure that design decisions support coherence and long-term sustainability [32].
4. **Small and simple interventions:** Human-centered design emphasizes iterative progress. Rather than rushing toward a final solution, designers should test small, simple interventions, learn from them, and refine the design over time [32].

This work adopts human-centered principles, with a particular emphasis on a user-centered design approach. Such an approach is crucial for addressing the challenge of communicating uncertainty in industrial contexts, where operators rely on clear, trustworthy information to make informed decisions. Furthermore, a user-centered approach requires an iterative process that prioritizes user feedback, maintains a clear focus, and incorporates continuous usability testing throughout the design cycle [33]. Central to this is developing a deep understanding of operator's current work practices, decision-making processes, and problem solving in critical situations. The research phase of this work will evaluate and investigate these aspects in the process industry.

3.3.1.1 User Experience (UX)

At its core, user experience describes how individuals interact with and perceive a product in real-world situations. And can be defined as all aspects of the relationship between a user and a product, including the systems and services connected to it [34]. It also involves the emotions, enjoyment, and satisfaction that users derive from engaging with a product, whether through using it or simply observing it. Importantly, while you cannot design a user experience, you can design for it [35].

3.3.1.2 User Interface (UI)

A User Interface (UI) is a part of any system designed to enable effective interaction with the end user. Within a system, the UI serves as an component, facilitating communication between the system and its users. It represents the main point, that the users engage with the system [36].

The goal of this work is to create guidelines, and graphical interfaces for communicating uncertainty in the process industry. Therefore User Interface (UI) is highly relevant to our project.

3.3.1.3 Participatory Design

Participatory design has emerged from the user-centered design tradition, emphasizing close collaboration between researchers and users throughout the design process. Rather than keeping users at a distance, researchers immerse themselves in users work contexts to understand their realities and challenges. Importantly, users are not merely consulted but actively participate in defining design goals and requirements based on their situated expertise [37].

The central goal of participatory design is to actively involve users throughout the design process, enabling them to contribute directly to the development of a product, system, or service. By engaging users at every stage, designers gain deeper insight into their needs and expectations, significantly increasing the likelihood that the final solution will be both relevant and effective for its intended audience [37].

Participatory design does not require the participating users to talk in technical terms, rather, it is achieved by several methods, such as interacting with prototypes and mockups [37].

As mentioned earlier in section 2.3, operators are experts users of the control rooms, with specific needs and specific work contexts, due to this, it is essential to include them in all steps of the design process, ensuring their feedback will be considered in the final design proposal.

3.4 Human-AI Interaction (HAII)

Human computer interaction has evolved rapidly as digital systems have shifted from traditional rule-based technologies to systems empowered by artificial intelligence (AI). This transition has given rise to a new area of study: Human-AI interaction

(HAII). The increased integration of AI into everyday tools has reshaped societal structures as well as numerous industries [38].

3.4.1 Human Centered Explainable AI (HCXAI)

As AI systems become more complex and increasingly integrated into everyday life, it becomes essential to ensure that their behavior is understandable to users. This need has driven the development of explainable AI (XAI), a set of methods aimed at making an AI systems reasoning transparent, comprehensible and explainable to users. This is of great importance in domains where incorrect decision-making can have major consequences, such as healthcare [39] . However, transparency alone does not guarantee that users will interpret or apply the information given correctly.

This gave rise to a new XAI framework, Human centered explainable AI (HCXAI). This framework does not focus solely on technical aspects, but highlights the importance of designing transparent AI systems based on the needs of users. Within this framework, the aim is to understand what kind of information a user needs in a given situation in order to make informed decisions, trust the system and interpret the information generated by the AI [5].

In the process industry, this is particularly important. Explanations must be adapted to the experience and skill level of operators, so that AI support becomes a practical aid rather than an obstacle to their work.

The human centered explainable AI framework identifies four core principles;

1. Operators should remain in control, and AI should act as a support for decision-making, not a replacement.
2. Explanations should be provided in multiple dimensions, for example, through technical details, natural language, visual representations, and clear communication of uncertainties.
3. AI systems should be adaptive and able to adapt to the situation and the user
4. AI systems should strengthen and complement the expertise, not replace it.

[5].

This work will primarily focus on the second principle, with particular emphasis on the communication of uncertainty.

3.4.2 Theories of Epistemic Uncertainty in AI Systems

Uncertainty has always been an inherent part of human knowledge making. The theory of uncertainty frames this not only as a technical issue, but as an epistemic limitation: uncertainty arises the moment we attempt to translate something from the real world, into a form that can be communicated. The gap between unity, i.e., the reality in its complexity, and something, i.e., our representation of that reality, is where uncertainty arises [40]. The epistemic challenge is particularly critical in AI systems. AI models inherently produce uncertain predictions due to

several aspects, such as incomplete data, model limitations, and the complexity of real world phenomena the system attempts to capture [2]. In contexts where AI predictions inform high stake decisions, failing to communicate this uncertainty can lead operators to overtrust AI outputs, making decisions out of incorrect predictions, that may result in harmful consequences.

4

Methodology

This chapter outlines the methodology used in the project, structured according to the Double Diamond design process. It focuses on why specific methods were chosen, while the following chapter (Process) describes how they were implemented in practice.

4.1 Design Process; Double Diamond

This work follows the design process of the Double Diamond (Figure 4.1).

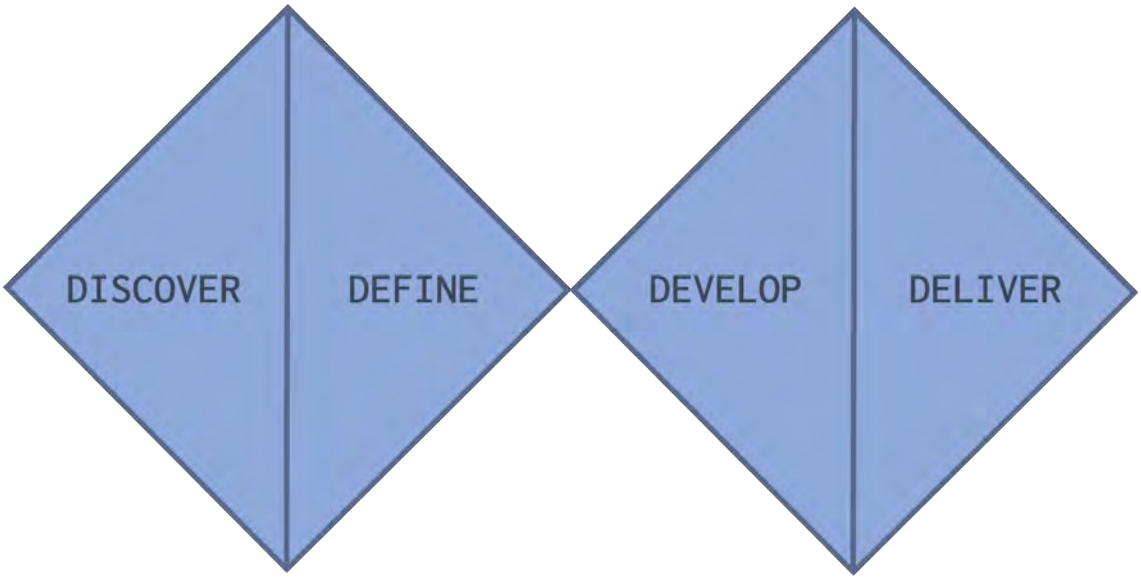


Figure 4.1: Double Diamond design process, adapted from the Design Council.

The Double Diamond design process is an iterative method that supports both divergent and convergent thinking through four distinct phases, visualized as two interconnected diamonds. The first phase, Discover, is about exploring and understanding the problem area in depth. The aim is to gather real facts, avoid early assumptions, and gain real insight into the user situation by involving those directly affected by the problem. The second phase, Define, involves structuring and analyzing the insights gained during the discovery process. Here, the problem area is formulated more clearly, often in a new way, based on the identified needs, patterns,

and challenges. The goal is to create a sharp and well-founded problem definition that can guide the design work further into the next diamond [41].

The second diamond consists of the Develop and Deliver phases. In the Develop phase, several possible solutions to the defined problem are explored and different concepts are generated that can address the needs of the users. This is followed by the Deliver phase, where the most promising ideas are prototyped, tested and refined to ensure that they actually work in practice and create value for the users [41].

The Double Diamond was selected for its systematic approach to first exploring the problem space through user research before developing solutions. Given the limited research on uncertainty communication in industrial contexts, this emphasis on deep problem understanding aligns well with the human-centered approach of the EXPLAIN project.

The methods employed in the Discover phase are presented first, followed by those used in the Define, Develop and Deliver phases.

4.2 Discover: Methods for Understanding the users Needs

This phase mainly focused on systematically exploring and understanding the problem domain, the target users, and their underlying needs and goals.

4.2.1 Literature Review

At the initial step of the discovery phase of the design process, an initial literature review was conducted, the review synthesizes the results and claims of previous studies on a topic and highlights the current state of knowledge in our research area. The literature review allowed us to deep-dive into existing research which enhanced our understanding of the topic. Performing a literature review offers several benefits: it can inspire new ideas to apply concepts in your own work, identify gaps or weaknesses in existing research, and highlight well-executed studies, helping to avoid unnecessary duplication of effort [42].

4.2.2 Semi-Structured Interviews

Semi structured interviews were conducted with operators working at a paper mill to gain deeper insights into their current work practices. Semi-structured interviews combine elements of both unstructured and structured formats [35], making them well-suited for exploratory research where understanding perceptions, unspoken shared agreements, and underlying intentions is essential [43]. Since the goal was to learn more about the users, their work, and their needs, incorporating unstructured elements helped create a conversational environment in which participants felt comfortable sharing their experiences, while the structured components provided consistency and supported later analysis [44]. Through these interviews, rich

insights were obtained into the operators everyday workflows, their attitudes toward existing systems, and their perspectives on integrating predictive technologies into their work.

4.2.2.1 Expert Interviews

Talking to experts in the exploratory phase is better for data gathering than other methods such as observations or quantitative surveys. Additionally, in expert interviews the interviewer and the interviewee share a common background on the subject which can increase the level of motivation of the expert [45]. Furthermore, involving experts in the design process offers several benefits. For example, they can provide insights from experimental research on micro-level processes and how decisions are made in real-world situations. Experts can also contribute context-specific knowledge that enhances the relevance of findings. When conducting interviews with experts, a semi-structured format is generally the most suitable approach [46]. The interviews performed in this work was conducted with paper mill operators, with varying levels of experience, see further demographics in section 5.1. The final analysis of the interviews can be found in section 5.3.1.

4.2.3 Observations

We conducted on-site observations at a paper mill to gain a deeper understanding of how the operators work, how the environment is structured, and how they navigate within the control room. These observations allowed us to see firsthand how they move between screens, communicate with each other to solve problems, and manage the systems in their daily workflows. The insights of the observations can be found in section 6.1.5. Observations, particularly in the early stages of design, can provide valuable insights into users context, tasks, and goals. By observing experts in their natural work environment, we can see how they perform their daily activities and interact with the tools they use. When conducting field observations, it is important to consider questions such as: who is using the technology, where it is being used, and how it is being applied [35], all these questions were considered during the on-site observation at the paper mill.

4.3 Define: Methods for Identifying the users Needs

The Define stage focused on clearly articulating the problem by analyzing insights gathered during the Discovery phase.

4.3.1 Thematic Analysis

A thematic analysis was conducted for identifying and organizing patterns in our dataset, as it is a flexible and fundamental method for qualitative analysis. This was done according to the six phases defined by Braun and Clarke [47]. These include (1) becoming familiar with the material, (2) systematically coding relevant parts of the data, (3) organizing these codes into preliminary themes, (4) reviewing

and refining the themes against both individual excerpts and the overall material, (5) clearly defining and naming each theme, and (6) compiling the analysis into a coherent and compelling report that links the themes to the research question [47]. The procedure of the thematic analysis is further described in section 5.3.1.

4.3.2 Design Guidelines

Design guidelines are typically context-dependent, practical recommendations grounded in empirical evidence and extensive experience. They provide direction during the design process, helping designers make informed decisions rather than prescribing fixed solutions. Because design contexts evolve, guidelines should be regularly reviewed and updated throughout the process. They are especially valuable in the early stages of design, when broad knowledge and direction are most beneficial. An effective set of design guidelines combines both general principles that apply across situations and more specific guidance tailored to particular contexts or problems [48]. In our work, the initial guidelines were developed through interviews, observations, and literature reviews, and were subsequently refined throughout the design process.

4.3.3 Persona

At the final stages of the define phase, a future oriented persona was developed. The purpose of creating a future persona was to establish a shared understanding of the anticipated user, their goals and their evolving work practices when interacting with AI driven decision support tools. Furthermore, personas are a goal-based design method where realistic and clear user archetypes are created based on information gathered through interviews and observations. Personas represent users goals, behaviors, and competencies, which can greatly help design teams make informed decisions about features and interactions. Personas are used throughout the whole design-process, as a lasting human reference [49].

4.3.4 Scenario

In addition, a future-oriented scenario was developed together with the persona. This scenario helped us further visualize the desired direction for the system, making it easier to align the team around a shared goal. Scenarios support design teams in exploring how users might interact with a product in the future, turning abstract ideas into tangible, easy-to-grasp narratives. The focus is placed on what users are able to do with the technology, rather than the technical details-while maintaining a strong user-centered perspective throughout. In this way, scenarios ensure that the entire design team has a clear and consistent vision of how the product, service, or system will be used in the future [44].

4.4 Develop and Deliver: Methods for Addressing the users Needs

For the develop and deliver phase different potential solutions were explored that can address the users needs, which were then prototyped, tested and refined.

4.4.1 Brainstorming

To kick off the idea generation of this work we used brainstorming. Brainstorming is a structured, group-based technique for promoting creativity in problem solving by creating an open environment [50]. The method encourage quantitative, broad solutions without any validation or criticism involved in the session. This approach helps groups unleash creativity by removing judgment and encouraging free thinking.

4.4.1.1 Crazy 8

The crazy 8 served as the method for the brainstorming to initiate the develop phase of our work. Crazy 8 is a brainstorming method designed to quickly generate a wide range of ideas [51]. The exercise is typically conducted in a group setting, where each participant is asked to sketch eight distinct ideas, one idea per minute. The goal is to encourage rapid thinking and ensure that each idea represents a new and different potential solution. Once all eight sketches are completed, participants present their ideas to the group, creating an opportunity for shared insights, discussion, and inspiration for subsequent design steps.

4.4.2 Prototyping

To further develop the ideas gathered from the brainstorming we started the prototyping. Prototyping is a key method in design processes, enabling teams to develop and validate both digital interfaces and physical products before final production. Prototypes can be created at different fidelity levels to match various stages of the design process. Low-fidelity prototypes, such as wireframes and sketches, are useful for early evaluation and testing basic concepts. High-fidelity prototypes include working functions and refined visual details, allowing for more realistic testing later in the process. This approach lets designers gather feedback and iterate efficiently at each stage [44].

4.4.3 Design Workshop

Design workshops are an effective co-design method for gathering creative input from stakeholders. They provide a structured and efficient way to collect feedback and involve stakeholders or participants in the design process. During the generative phase, workshops typically include exercises such as flexible modeling and other collaborative activities that support ideation and help verify potential design directions. For evaluation, participants are brought together to review developed concepts, offer feedback, and share insights that inform iteration and refinement of the design [44].

4.4.3.1 How Might We

The How Might We method is a design thinking tool used to transform defined challenges into open-ended, opportunity-focused questions [52]. By reframing our guidelines as How might we questions, teams create space for exploration, creativity, and multiple possible solutions.

The purpose of the design workshop were to evaluate the guidelines to see if designers can understand them and translate into design sketches. This approach helps shift from understanding the problem to generating ideas, acting as a crucial link between the Define phase and the Ideate phase of the design process.

The 'How Might We methodology' served as the starting point for our prototyping process, helping us broaden the range of potential design directions while still using the guidelines as an anchor to ensure that all ideas remained relevant and aligned with our objectives.

4.4.4 Evaluation

To test our prototype, we carried out an evaluation with operators. Evaluation research focuses on testing prototypes or interfaces with people who could potentially use the system. The purpose is to compare users expectations with the designed solution and assess whether it is useful, easy to use, and appealing. This type of research is iterative, meaning that the prototype is improved based on feedback from the users [44].

4.4.4.1 Think Aloud

Think aloud was used as a method in our final evaluation. Think-aloud is a well-established method in UX design that involves participants being asked to think aloud while performing a task. By verbalizing their thoughts, feelings, and actions during interaction with an interface, the design team gains valuable insights into, for example, points of frustration, useful features, and other aspects that affect the user experience. In this study, we used concurrent think-aloud, the most common variant of the method. This approach involves the participant continuously describing their spontaneous reactions and thoughts while navigating the interface. In this way, it becomes possible to capture both positive and negative experiences in real time. The method has major advantages when it comes to understanding users immediate experiences, but it can also feel unusual for some participants. Therefore, it is important for the test leader to remind them to continue verbalizing their thoughts throughout the task [44].

4.4.4.2 Interviews

Interviews was also decided to be one of the methods used in the evaluation of the prototype, and is a widely used method in evaluations. Interviews allow designers to gather deeper insights from participants, providing a qualitative approach to see areas of improvement and areas of satisfaction within a design. In the evaluation,

we decided to use a semi-structured approach, which allowed us to create a conversational environment where the participant felt comfortable in sharing their thoughts, ideas and feelings.

4.4.4.3 Van der Laan Acceptance Scale

The Van der Laan Acceptance Scale is a method used to collect quantitative data and is commonly applied in the evaluation of prototypes. The scale measures users acceptance of a technology through nine different statements, each rated on a five-point scale. It covers both perceived usefulness and the level of satisfaction, making it a valuable tool for assessing how well a system is perceived and received by users [53].

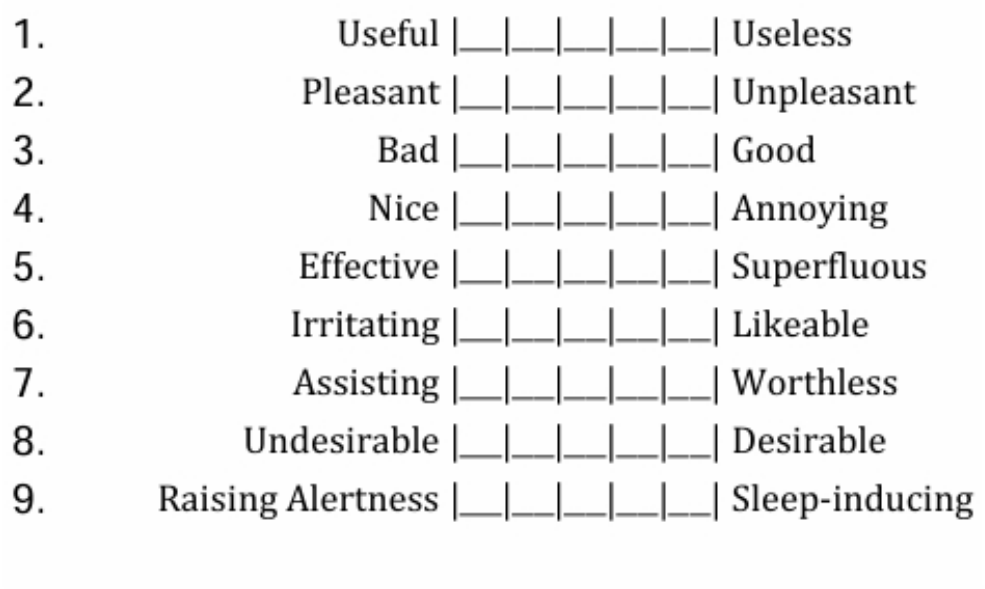


Figure 4.2: The 'Van der Laan acceptance scale'

4.4.5 Affinity Diagramming

Affinity diagramming is a method used to organize and interpret qualitative data, often collected through user research. It is inherently an inductive, bottomup process: researchers begin by grouping individual observations or insights into related clusters, which are then synthesized into broader, overarching themes. This approach helps reveal patterns and connections that might not be immediately apparent [44]. We used affinity diagramming to analyze and summarize the the findings from the interviews and think-aloud protocol.

5

Process

This chapter presents the execution of our project, which was structured according to the Double Diamond design process, as explained in the previous chapter. The chapter provides an overview of the four phases of the process and describes the methods, activities, and decisions undertaken in each stage (See figure 5.1). The process began with interviews, observations and a literature review, which together formed the basis for an initial set of design guidelines. These were then brought into a workshop with UX designers and engineers, where the goal was to examine whether the guidelines were understandable and could be translated into concrete design suggestions - resulting in a revised set of guidelines. Finally, the guidelines were implemented in a prototype and evaluated with real operators, grounding them in actual use and informing the final iteration of both the prototype and the guidelines. See figure 5.2 for an overview of the process.

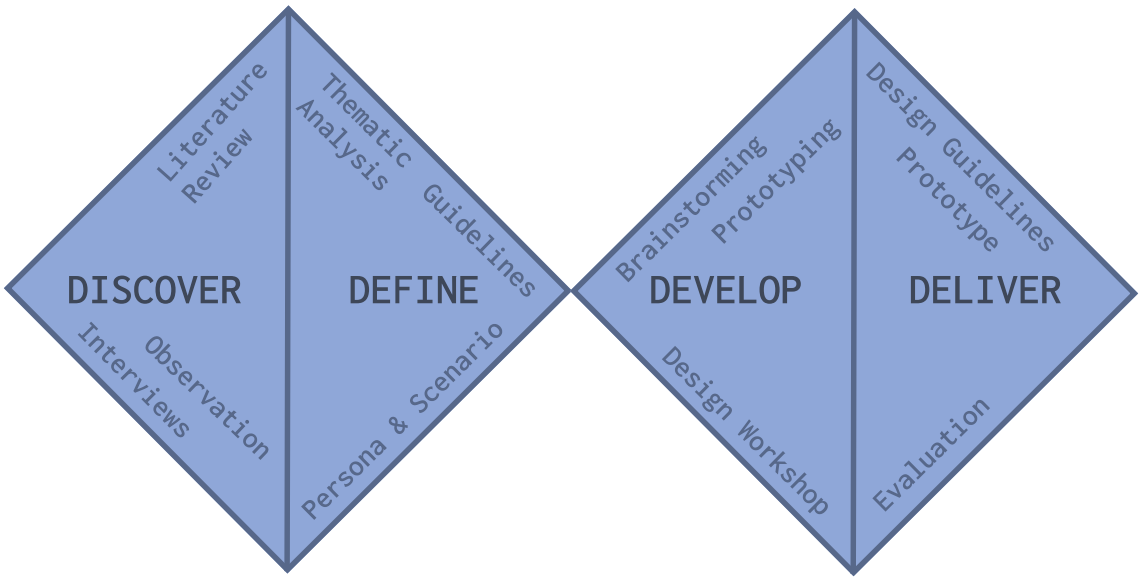


Figure 5.1: Double Diamond design process, including methods in each phase

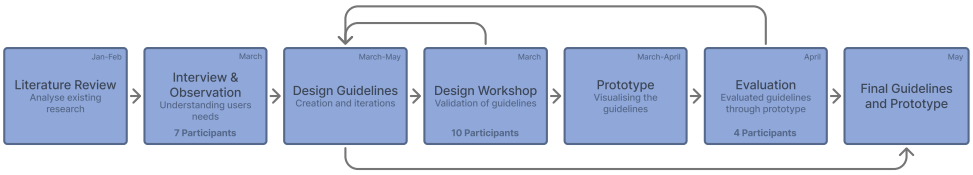


Figure 5.2: The iterative design process across January-May, incorporating three participatory phases: interviews and observation, a design workshop, and evaluation.

5.1 Discover

The purpose of the discover phase was to build a solid understanding of the users and their needs. This was carried out through a literature review, which provided us with clarifications of the current research and the current knowledge gaps. Additionally, interviews were conducted with operators from the paper industry, with the aim to gain further knowledge about current work practices, their approaches to uncertainty in their work, and attitudes towards predictive systems.

5.1.1 Initial Literature Review

The first phase of this work consisted of a comprehensive literature review. This was carried out by examining existing research related to the core themes of our study, including current approaches to uncertainty visualization, the role and practices of industrial operators, and the integration of AI into the process industry. In parallel, we also explored theoretical frameworks relevant to our main focus - designing for uncertainty. These included User-centered design, Explainable AI, mental models, decision-making processes, and related design and cognition theories.

In addition, the EXPLAIN project conducted by ABB Corporate Research offered valuable insights into explainable AI methodologies specifically within the process industry context. This project significantly contributed to our understanding of how AIdriven decision support can be made transparent and interpretable for operators. Overall, the literature review provided a solid foundation for our work. It clarified the current state of research, highlighted existing knowledge gaps, and helped us position our contribution within the broader academic and industrial landscape.

The literature review can be found in chapter 2 and 3 of this report.

5.2 Recruitment of Participants

Participants were recruited using convenience sampling. One of the researchers had direct contact with an operator at Holmen Paper Mill, who facilitated contact with a process engineer. The process engineer then supported the logistics and scheduling of both the initial interviews and the final evaluation.

For the initial interviews, one shift team was included, and for the final evaluation a different shift team was selected. In both cases, participants were chosen based on which employees were working on the days the data collection sessions took place.

5.2.1 Semi-structured Interviews

The operators currently do not use any predictive systems. Therefore, for the semi-structured interviews, the aim was to gain a solid understanding of their current work practices, approaches to handling uncertainty, attitudes towards visualizations in existing systems, and perspectives on predictive systems.

The interviews were conducted online via Microsoft Teams with operators at Holmen Braviken paper mill working within the TMP bleaching process. In total, 6 operators were interviewed. Although the number of participants was relatively low, we prioritized gaining deep understanding over breadth of insights [54]. The in-depth nature of the interviews provided sufficient information about their work practices, approaches to uncertainty, perspectives on current data visualizations, and attitudes towards predictive systems - including what they would need to trust such systems.

5.2.1.1 Formulation of Interview Questions

Before conducting the interviews, we worked on formulating relevant and comprehensive interview questions to ensure that the information we needed would be captured at this stage. The preparation process included a thorough review of the operators work situation as well as a clear articulation of our use case - the TMP bleaching process. Our aim was to enter the interviews with a foundational understanding of their daily work, enabling us to ask more precise follow-up questions based on the participants responses.

The development of the interview questions was an iterative process in which we placed great emphasis on understanding the purpose of each question. The questions were divided into five sections, which are described below along with their respective objectives. The interview questions translated into English can be found in Appendix A.1.

1. Demographics: The first part of the interview focused on questions about the operators level of experience, educational background, age, and gender identification. The purpose was both to create a basic profile of the participants and to enable the compilation of demographic data.
2. Work Context: The second section addressed the operators work context. Here, we asked questions about what a typical workday looks like, which

systems they use on a daily basis, and how these systems integrate into their workflow.

3. Uncertainty: In the third section, we explored the operators experiences of uncertainty in their work. The aim was to understand their current levels of uncertainty, where uncertainty arises, how they manage it, and how decisions are made under uncertain conditions.

4. Distributed Control System (DCS): The fourth section addressed the current work-practices and attitudes towards the distributed control system. Here, we wanted to get information on how the operators navigate in the system, which visualization techniques they prefer, and if they find any visualizations difficult to interpret.

5. The AI concept: The fifth and final section introduced the concept of predictive systems and investigated the operators attitudes toward AI-based solutions. We focused on what they would need in order to trust such a system, which features would be most valuable to their work, and how a predictive system could potentially streamline and support their current tasks.

Through the formulation of the interview questions, we aimed to gain a further understanding of operators current work practices, their approach to uncertainty and their attitudes towards predictive systems.

5.2.1.2 Interview Procedure

Before the interviews, participants signed a consent form that provided information about the research project, their rights, and the types of personal data that would be collected. Due to specific requirements from Holmen, the interviews could not be recorded. This restriction was also clearly stated in the consent form (See Appendix A.4).

Each interview began with an introduction to the project and its purpose, followed by a reminder of the participants rights. Participants then answered demographic questions before starting the interview.

The semi-structured interviews were guided by predefined questions, with additional follow-up questions tailored to the participants responses. The questions were also adapted depending; for example, if a participant described a situation where they experienced difficulty making decisions, we referenced the same example in subsequent AI-related questions to help them provide more informed answers.

During the interviews we had clearly defined roles, one interviewer asked all the questions, while the other interviewer took detailed notes of the participants' responses. Each interview lasted approximately 20-30 minutes. In total, six operators were interviewed. All participants were male, with varying levels of experience working as operators. This information is summarized in table 5.1.

Participant	Role	Experience	Education	Age	Gender
1	Operator	5-10 years	High School	30-39	Male
2	Operator	20+ years	4-year High School	40-59	Male
3	Operator	1-3 years	High School	18-29	Male
4	Operator	1-3 years	High School	30-39	Male
5	Operator	20+ years	High School	60-70	Male
6	Operator	20+ years	High School	40-59	Male

Table 5.1: Demographic information of participants from the initial interviews at Holmen

5.2.2 Observations

In addition to the initial interviews, a visit to Södra Cell pulp mill took place, where operators working in a control room responsible for the bleaching process were observed. During the observations, operators were followed for approximately half a day, during which additional questions were asked regarding their tasks, needs, decision-making processes, and situations involving uncertainty. Since AI was already used in the bleaching process at Södra, questions were also posed about attitudes toward the technology, as well as trust in and uncertainty surrounding AI. During the observation, we primarily interacted with a male operator (aged 29, two years of experience) responsible for monitoring and managing the bleaching stage. His participation was incidental rather than deliberate, as he was the operator available during our visit.

5.2.3 Secondary Literature Review

Since the operators from the initial interviews at Holmen paper mill did not currently use a predictive AI system, a secondary literature review was conducted to complement the interview findings. As research on designing for uncertainty in predictive systems within the process industry remains limited, the review extended to other domains, including healthcare, gaming, and geospatial visualizations, to explore whether insights and recommendations from those fields could be applicable in an industrial context. By integrating these findings with the perspectives gathered from the interviews, we developed a set of design guidelines grounded in both existing research and practitioner experience.

Findings from the secondary literature review can be found in section 2.4.

5.3 Define

The purpose of the Define phase was to narrow and structure the insights gathered during the Discover phase. This was achieved by analyzing the interview and observational data and formulating design guidelines based on insights from the interviews, observations and the literature review. Additionally, personas and scenarios were created to illustrate key user needs and contextualize the guidelines for

subsequent ideation and prototyping.

5.3.1 Thematic Analysis

To analyse the material from our interviews and observations, we conducted a thematic analysis according to the six steps defined by Braun and Clarke [47]. See figure 5.3 for an overview of the process.

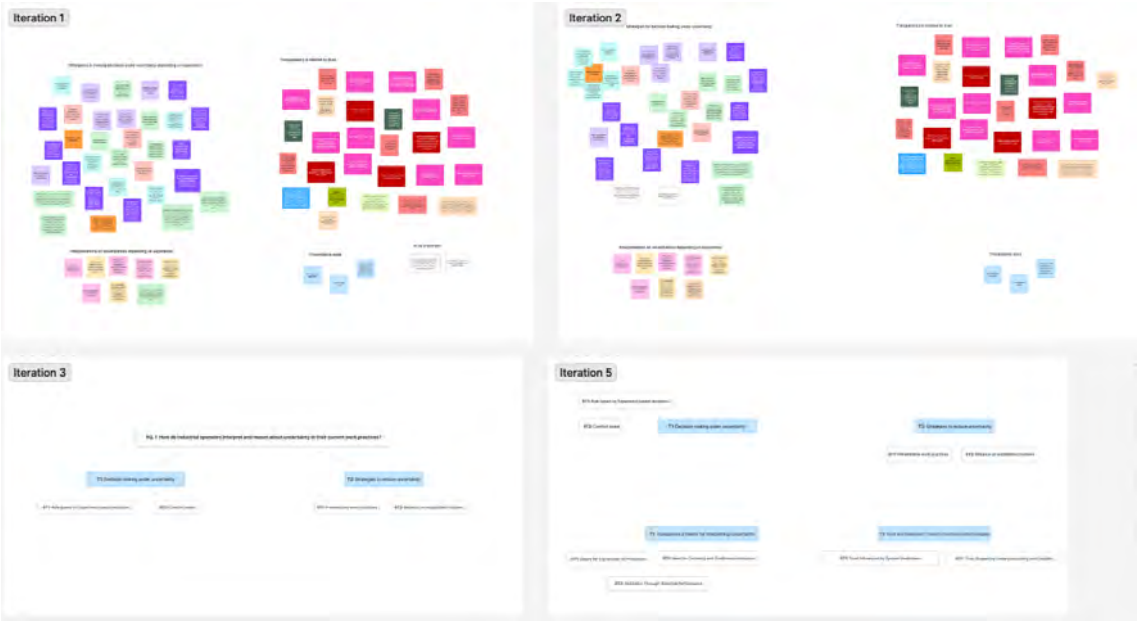


Figure 5.3: Overview of the process of the thematic analysis

Since the interviews were not recorded, we began the process by reviewing all the notes from the interviews and observations. Where necessary, we made clean-up edits in the form of clarifications and corrections of spelling errors. Once we had become thoroughly familiar with the material, we began to structure it in FigJam, where each interview note was placed together with the associated interview question. This gave us a clear overview of the collected material and formed the basis for our coding process.

The coding was carried out jointly through continuous discussions about how we best wanted to categorize the data. Once all the material was reviewed and coded, we did another reading to ensure that all relevant information had been captured. In some cases, interview responses were double-coded when we judged that they touched on several significant areas.

We then began the work of identifying themes from the coded data. This was done by merging related codes into more comprehensive and coherent themes. We initially named themes in a way that reflected their content, but as the process was iterative, the themes were reformulated several times during the analysis. Once the main themes were clearly defined, we also identified sub-themes to create a clearer and more nuanced structure that fairly represented the content and meaning of the data.

Once we agreed upon the fact that all relevant data from the dataset were represented in one of our themes, we continued to work with the naming to further clarify the titles of the themes.

The final themes extracted from the dataset were the following: *Decision making under uncertainty, Strategies to reduce uncertainty, Transparency Needs for Interpreting Uncertainty* and *Trust and Skepticism Toward Uncertain System Outputs.* These themes will be further described in section 6.1.

In addition to the themes generated through the thematic analysis, some interview and observational insights that were not included as part of the final themes were still considered relevant when formulating the design guidelines and important to have in mind. These insights did not directly answer the research questions and therefore were not incorporated into the thematic structure. To maintain transparency, these complementary insights are reported separately and are not presented as part of the formal thematic analysis. These are reported in section 6.1.5.

5.3.2 Defining the Design Guidelines

To establish the design guidelines, we began by systematically reviewing the literature and identifying insights relevant to uncertainty communication in predictive systems. These findings were then compared against the insights derived from the thematic analysis of our interviews, and where the two sources aligned, we consolidated them into a design guideline. This approach ensured that each guideline was grounded in both existing research and the lived experiences of operators.

One exception was the guideline *the design should communicate both positive and negative outcomes.* This guideline emerged from the literature alone, as no corresponding insight was raised by the operators during the interviews. However, given the influence of framing effects on decision-making under uncertainty (see section 2.4.1), we judged it important to include and to test whether it would resonate with operators in practice.

The guidelines, including descriptions, are presented in Table 5.2. The full final guidelines with descriptions and rationale are presented in Section 6.5.

Nr.	Guideline	Description
1	The design should be based on users context, experience, needs, and tasks	People interpret uncertainty differently based on their background and expertise. Experienced operators use established mental models, while less experienced ones rely on guidance and routines. Tailoring visualizations to these different needs ensures uncertainty is communicated meaningfully and supports better decisions (see [16], [17], [20], [23], 6.1.1).
2	The design should aim to reduce cognitive load and clutter	Human short-term memory is limited, making information overload a risk in complex decision environments. Beyond just limiting information, design should provide clear reference points and contextual cues (see [9], [20], 6.1.5.2).
3	The design should provide certainty and confidence indicators	Clear confidence indicators help users understand prediction reliability and make informed decisions rather than blindly trusting output. Operators emphasized this as fundamental for assessing reliability of the predictions into their decision-making (see [2], [5], [14], [15], [16], [18], 6.1.3.2).
4	The design should provide clear explanations of predictions	users need to understand how the system reasons, main parameters it uses, and historical accuracy. This transparency enables operators to verify the system's logic aligns with their own understanding, building trust and supporting well-grounded decisions (see [20], 6.1.3.1).
5	The design should provide historical performance	Operators expressed that the key to trust the predictions is by presenting history with context about the conditions under which predictions were made, so operators can judge relevance. Showing how similar situations were handled previously also builds confidence (see [20], 6.1.3.3).
6	The design should communicate both positive and negative outcomes	Cognitive biases such as the framing effect makes people respond differently to the same information depending on whether it's presented positively or negatively. Displaying both perspectives supports more balanced, informed interpretation of uncertainty (see [16]). This guideline emerged only from literature.

Table 5.2: First iteration of design guidelines for uncertainty communication based on insights from literature and initial interviews.

5.3.3 Persona and Scenario

At the final stages of the define phase of the Double Diamond, a future-oriented persona and scenario were developed. While grounded in insights from the operator interviews, these artefacts do not describe the current state of work at the mill. Instead, they represent an envisioned future context in which predictive AI capabilities are integrated into the control room environment.

The purpose of creating a future persona and scenario was to establish a shared understanding of the anticipated user, their goals, and their evolving work practices when interacting with AI-driven decision-support tools. By situating the persona in a future workflow, we could explore how operators may collaborate with predictive systems, how their decision-making processes might shift, and what forms of support they would need in situations shaped by real-time forecasts and uncertainty information. The persona and scenario will be presented in the following sections.

5.3.3.1 Future Persona

Figure 5.4: User persona: Per, a 45-year old paper mill operator.

Per has worked as an operator in the paper mill for 15 years. Before entering the industry, he completed high school. His role involves shift work, alternating between day and night shifts, although his core responsibilities remain largely unchanged regardless of the time of day.

Pers overarching goal is to keep the production process running smoothly. He primarily monitors the process from the control room through the distributed control system, where he typically works across four screens. On certain days, he is responsible for the entire production line, requiring him to oversee up to eight screens. Each screen presents different views corresponding to various stages of the pulp production process, and over the years, Per has established his own routines for navigating and using these views efficiently.

Per increasingly incorporates AI-based predictive tools into his daily routines. While he continues to rely on his experience and established navigation patterns across multiple screens, the predictive system becomes an important complement to his own judgment. Instead of only reacting to deviations, Per uses realtime forecasts and uncertainty information to anticipate changes earlier and prepare actions in advance. This proactive mindset helps maintain stable and efficient production conditions.

When problems arise, Per draws heavily on the strong sense of companionship within his shift team. He often discusses smaller decisions with his colleagues, valuing their experience and collective problem-solving ability. For more significant decisions, he maintains close communication with the shift leader, who ultimately holds the formal responsibility for major operational choices. This collaborative decision-making structure contributes both to effective problem-solving and to the supportive environment Per experiences in his daily work.

5.3.3.2 Future Scenario

Per is working a day shift in the control room. Production is running smoothly, and soon the process will transition to producing pulp with higher brightness. This means the pulp will undergo dithionite dosing to achieve the required brightness level. One of Pers most critical tasks is determining when the bleaching process should begin.

In the past, Per relied on static recipes that indicated roughly when bleaching should start. Although useful as guidelines, these recipes did not account for real-time variations in the process and were therefore not always accurate. As a result, operators typically choose to start bleaching slightly earlier than necessary because the economic loss of bleaching too early is much smaller than the consequences of bleaching too late.

However, this strategy often led to excessive use of dithionite, a chemical that represents a significant cost for the mill.

Today, Per works with an AI-based prediction system that incorporates all relevant process parameters in real time and forecasts when the brightness is expected to increase. The system serves as a decision-support tool during every transition and has become a natural part of Pers workflow. The AI system clearly communicates all the information Per needs in order to make informed decisions, such as historical prediction accuracy, the model parameters and the uncertainty and confidence of the predictions.

By using the prediction as a reference, Per can make more well-grounded and precise decisions about when the shift team should initiate bleaching. This enables him to:

1. Deliver light quality on time,
2. Avoid unnecessarily early bleaching starts
3. Minimize chemical waste

4. Ensure a more predictable, stable, and costefficient process.

The new system does not replace Pers expertise, but it significantly enhances his efficiency and reduces uncertainty in one of the processs most critical moments.

5.4 Develop and Deliver

In the development and deliver phase, we built on the established guidelines by brainstorming and refining ideas. These guidelines were validated through a design workshop, allowing designers and engineers to test their applicability in practice. We then iteratively developed and evaluated prototypes, using an interactive prototype to assess the guidelines, and refined both the prototype and the guidelines based on our findings.

5.4.1 Brainstorming

When we had identified and defined all the guidelines we started the development phase by using brainstorming methods to initially kick-off the idea generation of this work.

5.4.1.1 Crazy 8

To kick-off the idea generation, the Crazy 8 brainstorming method were used. We began by establishing the rules: each group member was to generate eight ideas per guideline within eight minutes. We repeated this process for all six of our guidelines. In total, we produced 93 ideas. Although we did not manage to create eight ideas for every single guideline, the overall outcome still provided us with a broad and diverse pool of concepts to work with. See figure 5.5 for an overview of sketches brainstormed.

Figure 5.5: Overview of brainstormed sketches during crazy 8

Once we finished the brainstorming session, we compiled all of our ideas in Figma and organized them according to the guideline they belonged to. This gave us a clear visual overview of our idea landscape and made the upcoming voting process much more manageable.

The next section will describe the dotvoting technique we used to narrow down the number of ideas.

5.4.1.2 Dot-voting

To navigate and prioritize the large number of ideas generated during brainstorming, we used dot voting as a method to narrow them down. Each of us received four votes per guideline. This allowed us not only to evaluate which ideas felt most relevant in relation to each specific guideline, but also to highlight patterns. Several of our ideas were very similar, and dot voting helped us identify where our thinking overlapped.

By visually seeing which ideas accumulated the most dots, it became easier to filter out weaker or duplicate concepts and move forward with the ideas that held the strongest potential. See overview of dot-voting in figure 5.6.

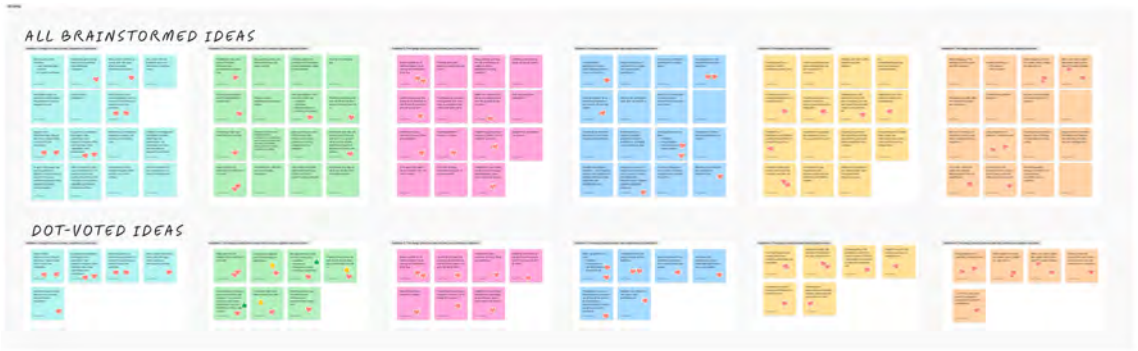


Figure 5.6: Overview of dot-voting

5.4.2 Low-fidelity Prototyping

After voting on ideas for each guideline, we summarized them and discussed the advantages and disadvantages of each concept. In the end, we had around two ideas per guideline, which we mapped into two separate prototypes. Both prototypes were designed to follow our guidelines, but they differ in terms of design and visualization.

We then began paper-sketching the two prototypes by hand. And then translated these design sketches into wire-frames in Figma. See initial sketches in figure 5.7 and 5.8.

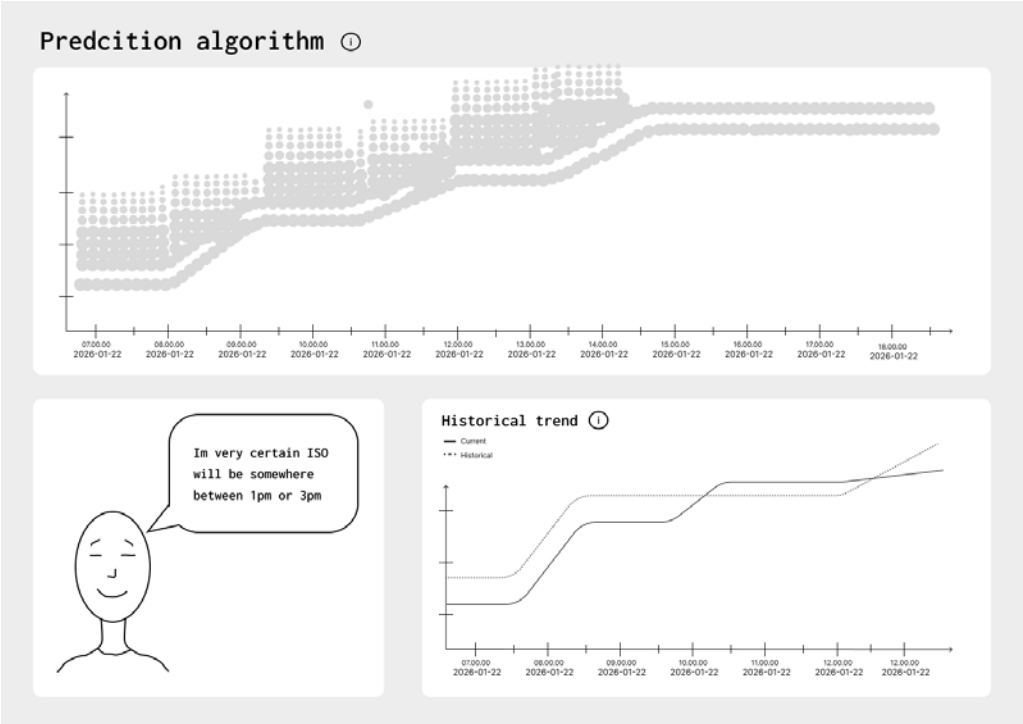


Figure 5.7: Initial prototype: alternative 1

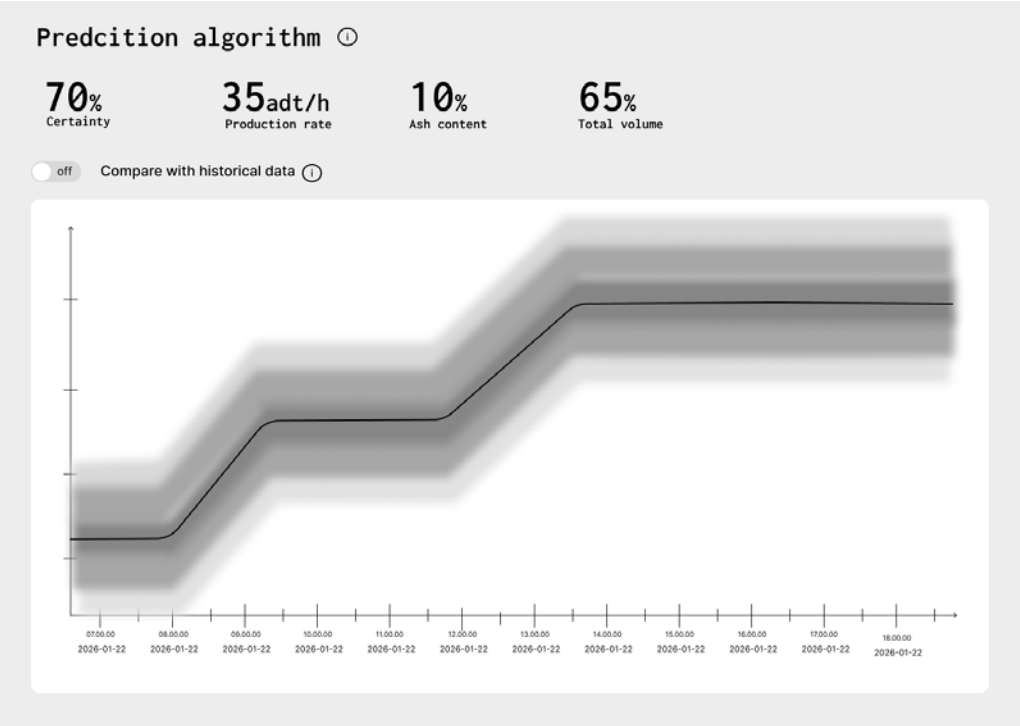


Figure 5.8: Initial prototype: alternative 2

5.4.3 Design Workshop - 'How Might We'

To evaluate the guidelines, we examined whether designers and engineers could understand and translate them into concrete design solutions or requirement specifications. To avoid biasing participants' designs, we deliberately chose not to share the initial sketches generated during the low-fidelity prototyping phase.

The evaluation was conducted as a design workshop with colleagues at ABB Corporate Research, including UX researchers, interaction design students, and engineering students. In total, two workshops were held with 10 participants: five engineering students, two interaction design students, and three UX designers and researchers. Both interaction designers and engineers were included to assess whether the guidelines were clear and actionable across different professional perspectives.

The workshops were structured around "How Might We" questions, which provided a framework for participants to interpret each guideline and generate design responses, allowing us to assess whether the guidelines were sufficiently clear and actionable to guide real design work.

The design workshop agenda included a:

1. Introduction to the project and use case
2. Presentation of persona, scenario and guidelines.
3. Warm up 'how might we' exercise.
4. Main How might we ideation session

5. Open discussion and reflection.

To help participants familiarize themselves with the use case, potential users, and the future scenario in which our design will operate, we began by clearly outlining the projects aim: to investigate how to design AI systems in contexts where outputs are uncertain. We introduced the use case - specifically the bleaching process - and encouraged questions to ensure shared understanding. Participants also received the persona, scenario, and guidelines to review independently, allowing them to build a deeper understanding of the users, context, and design considerations.

After the introduction, we facilitated a short warm-up exercise to ease participants into ideation. Using the prompt How might we design the worlds best ice cream?, we asked participants to intentionally generate the worst ideas they could think of, helping them loosen up and think more creatively.

We then continued with the main set of How might we questions. Each prompt began with How might we and was directly connected to one of the proposed design guideline. See figure 5.9 for how might we questions.

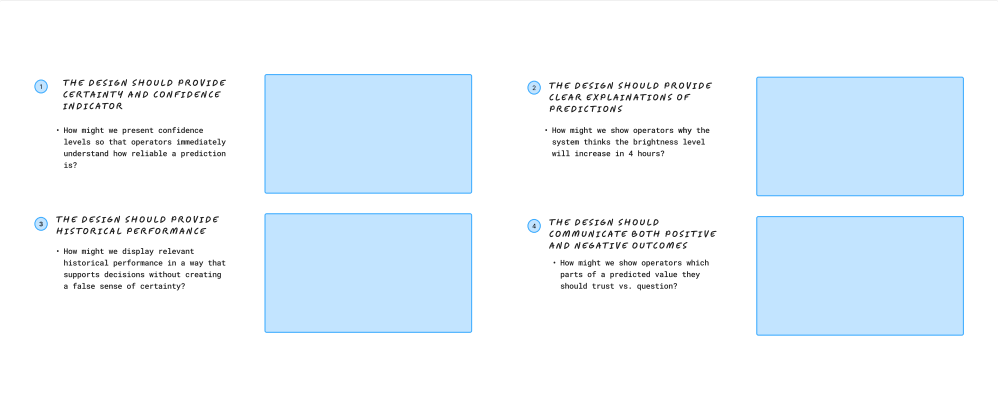


Figure 5.9: How might we questions linked to each guideline

Due to time constraints, we chose to evaluate only four of the six guidelines during the workshops. The first two *The design should be based on users context, experience, needs and tasks* and *The design should aim to reduce cognitive load and clutter* were excluded, as their general nature made them less suited to the "How Might We" format. Unlike the remaining guidelines, which relate to concrete interface decisions, these two concern broader design guidelines are difficult to translate into specific ideation suggestions. We therefore prioritized the four guidelines that we considered most actionable in this context.

To conclude the workshop, each participant shared their thoughts and ideas related to the guidelines. This closing discussion gave us the opportunity to ask followup questions, clarify points of confusion, and gain additional insights into how participants interpreted the guidelines.

After the How might we workshop, all of the ideas were reviewed, along with the notes taken during the session. Overall, most participants expressed that the guide-

lines and the How might we questions were clear and easy to translate into sketches. However, several participants noted that the final guideline was more difficult to interpret and harder to fully understand.

5.4.4 Iteration of Design Guidelines

Based on the insights from the workshop (see Section 6.2), we decided to remove the guideline: *The design should communicate both positive and negative outcomes.* Most participants found it difficult to understand and struggled to translate it into concrete design solutions or requirement specifications, which was the main thing we were trying to evaluate in the workshop. This was further reinforced by the fact that the guideline had not emerged from the operator interviews and was only identified in the literature, meaning it lacked grounding in actual operator needs. Together, these two reasons led us to remove it from the final set of guidelines.

5.4.5 High-fidelity Prototyping

Following the initial brainstorming and workshop sessions, the ideas generated collaboratively were consolidated and subjected to a second round of voting to determine which elements should be included in the final design. The prototyping was carried out in Figma [55], which allowed for iterative visualisation and refinement of the selected concepts. Throughout the prototyping process, all design decisions were informed by the established guidelines. The following subsections describe how the guidelines were addressed in the design. See prototype used in the evaluation in figure 5.10 and 5.11.

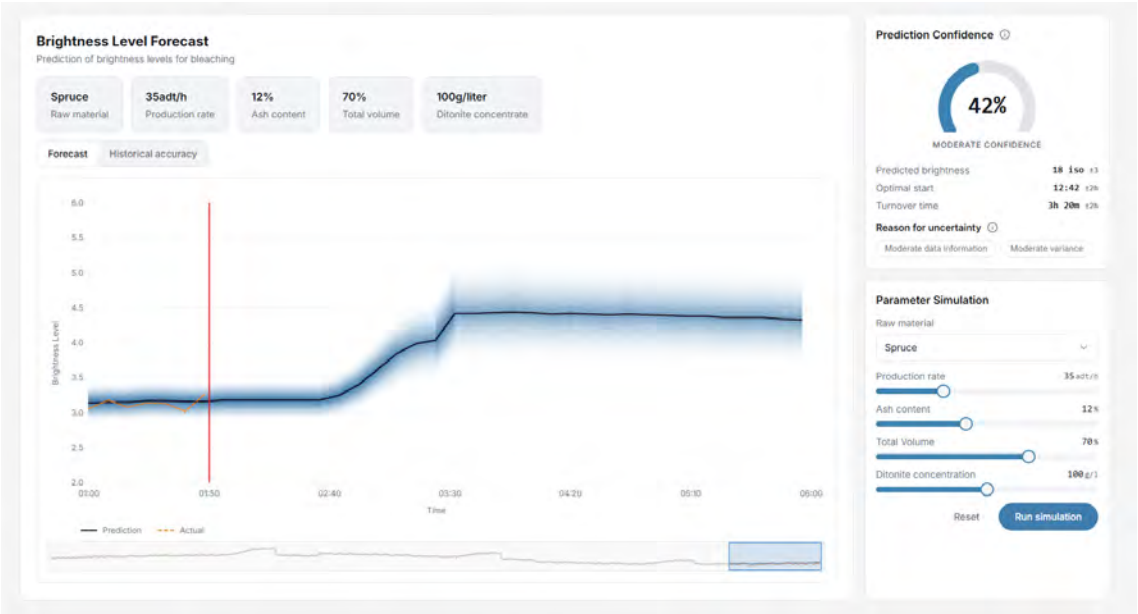


Figure 5.10: Main view of prototype used in the evaluation

Figure 5.11: Historical accuracy view of prototype used in the evaluation

5.4.5.1 Designing for the users Context, Experience, Needs, and Tasks

A key priority during prototyping was grounding the design in a realistic use case, so that the subsequent evaluation would yield meaningful and valid feedback. The initial interviews revealed considerable variation in operator experience. Less experienced operators expressed a need for detailed explanations of the system's reasoning, whereas more experienced operators relied primarily on their domain knowledge and required less contextual information. To accommodate both ends of this, the interface was designed to show only the most essential information by default, while making deeper explanations available on demand, for instance, through the contextual info icons described in Section 5.4.5.4.

5.4.5.2 Designing to Reduce Cognitive Load and Visual Clutter

When designing to reduce cognitive load and visual clutter, we focused on eliminating both unnecessary information and distracting visual elements, such as grid lines and chart borders.

5.4.5.3 Designing Confidence and Uncertainty Indicators

Several visualization approaches were explored and evaluated to communicate prediction confidence before settling on a design that combines a numerical percentage with a gauge. This dual representation was chosen to ensure that operators could interpret confidence both precisely and at a glance.

For representing predictive uncertainty in the time series plot, a blur-based visualization was selected over alternatives such as circular glyphs. The blur approach was preferred both for its support in the literature [20] and because it was judged to convey uncertainty more intuitively. Matplotlib [56] was selected for visualizing the

plot, as it offered greater programmatic control over blur rendering than Figma's [55] native capabilities allowed. Given that this required writing custom visualization code beyond the scope of the project's primary development work, Claude AI [57] was used to assist in drafting and refining the implementation, with the final code reviewed and validated to produce the intended visual output.

5.4.5.4 Designing Clear Explanations of Predictions

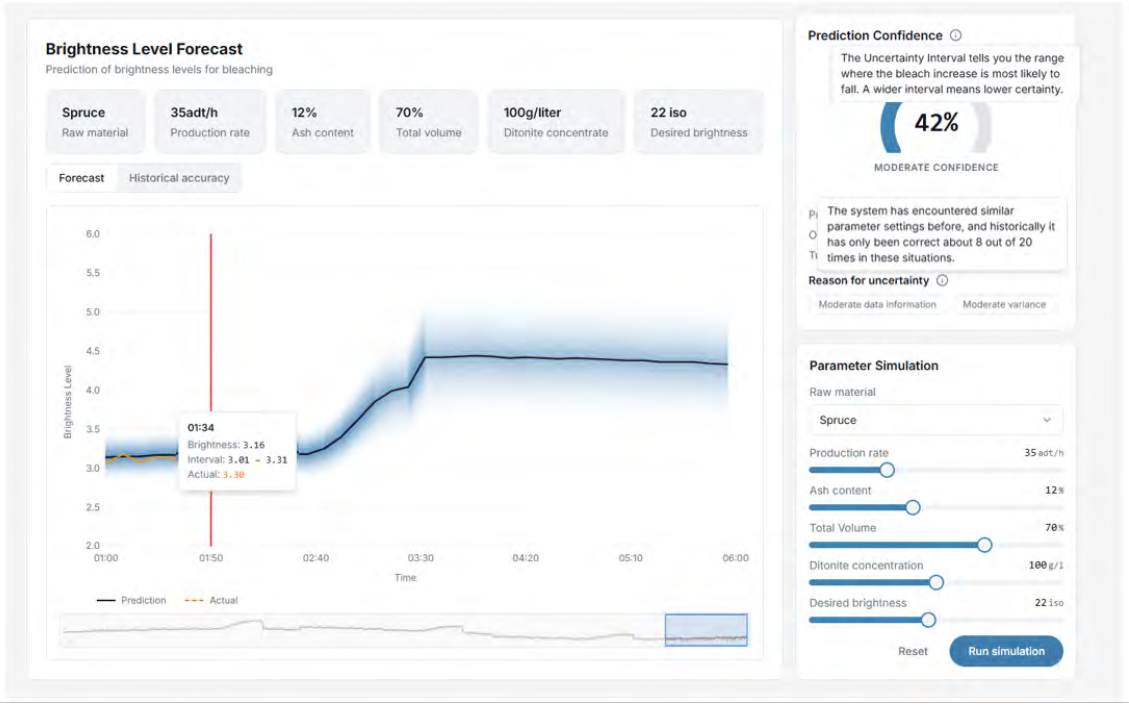


Figure 5.12: Prototype with hover over info

To support operator trust and transparency, the interface was designed to communicate which input parameters the model had considered when generating a prediction. These parameters are displayed as cards positioned in the top of the screen. This decision was motivated directly by findings from the initial interviews, in which operators stated that they would not trust an AI system unless it could demonstrate that it reasoned about the same parameters they themselves would consider. Operators wanted to be able to assess whether the AI's reasoning aligned with their own expectations under uncertainty.

To keep the interface uncluttered while still providing depth of explanation, supplementary information was made accessible through contextual info icons. These icons appear alongside both the prediction confidence indicator and the uncertainty indicators, and when hovered, they provide explanations of what the confidence means and why the model is uncertain in a given situation. This allowed detailed explanations to be available without overwhelming the primary view (see figure 5.12).

5.4.5.5 Designing the Historical Accuracy

Several approaches to presenting historical model performance were considered, including an adaptive screen layout intended to consolidate all information into a single view. However, this approach was ultimately rejected on the grounds that it would place too much information on a single screen. Historical accuracy was therefore moved to a dedicated secondary view. In addition to accuracy metrics, the design includes an indicator of model accuracy, allowing operators to identify cases where, for instance, a prediction may have been correct even when the associated confidence level was low - or vice versa.

5.4.5.6 Designing the Parameter Simulation

Although a parameter simulation feature was not part of the original design guidelines, it emerged as a recurring interest in both the initial interviews and the design workshop (See section 6.1 and 6.3). Operators expressed a desire to explore different scenarios, for example, to understand how a change in raw material or total volume would affect the brightness forecast. This feature was therefore included in the prototype to gather evaluative feedback on its utility and usability. The simulation panel allows operators to adjust the primary input parameters driving the prediction and observe the effect on the forecast output.

5.4.5.7 Making the Design Interactive

Once the designs were complete, a decision was made to transform them into a fully interactive prototype for the final evaluation using Lovable, an AI-assisted tool capable of transforming design sketches into functional, scalable web prototypes [58]. This was motivated by that certain features, particularly the parameter simulation, would be difficult to convey meaningfully through static screens alone or through prototyping in Figma, and that a more realistic experience would lead to higher-quality evaluative feedback.

The interactive prototype was built using Lovable [58]. The Figma designs were provided to Lovable alongside a detailed prompt describing the project context, the intended behavior of each interactive element, and the goals of the evaluation. To maximise the clarity and effectiveness of this prompt, it was drafted collaboratively with Claude AI [57].

The initial interactive prototype produced by Lovable closely matched the intended design, though several elements required refinement. These were addressed through targeted follow-up prompts accompanied by the relevant design sketches. One notable challenge was the time series plot, which did not render as intended from the design sketches from Figma alone; this was resolved by giving Lovable the code developed earlier to create the plot with Matplotlib [56], which enabled the desired visualisation to be reproduced in the interactive prototype.

5.4.6 Evaluation of Prototype

The prototype was evaluated using a combination of methods, including a think-aloud session , the Van der Laan Acceptance Scale, and semistructured interview questions. The purpose of this evaluation was to examine how well the proposed guidelines reflect real operator needs, assessed through a prototype designed according to those guidelines. In addition, we sought to understand operators trust in such a system and explore how communicating uncertainty supports them in making informed decisions.

5.4.6.1 Planning of Evaluation

We began planning the evaluation to be conducted with operators at the Holmen Braviken paper mill. Since we knew that only four operators would be available, we focused primarily on gathering qualitative insights, as the sample size was too small to generate meaningful quantitative results. We therefore selected semistructured interviews as our main evaluation method, complemented with a think aloud session and the Van der Laan Acceptance Scale to capture some quantitative indications. Although the quantitative data would not be statistically significant, it could still provide supportive insights and help contextualize the qualitative findings.

The questions were carefully formulated to align with our research questions, ensuring that the evaluation remained focused on our aim and served as a solid foundation for the continued work in the process. A total of nine questions were developed, and since the format was semi-structured, follow-up questions were also posed to all participants depending on their responses. See the interview questions translated into english in Appendix A.2.

Additionally, the Van der Laan acceptance scale was used to gather complementary quantitative data. The original scale consists of nine questions answered on a five-point scale ranging from -2 to +2, with each question contributing to one of two subscales: usability and satisfaction. To ensure we captured insights directly relevant to our research questions, we chose to supplement the scale with three additional questions: Reliable/ Unreliable, Easy to understand/ hard to understand, Clear/ Unclear. These three additional questions were included in a third sub-scale, comprehensibility and trustworthiness. See the complete scale translated to english in Appendix A.3.

5.4.6.2 Evaluation Procedure

The evaluation was conducted in person with operators at the Holmen Braviken paper mill. Before beginning the main session, each participant was asked to read and sign the informed consent form. We also asked the participants questions about their demographics (see table 5.3). Once consent was obtained, we initiated the evaluation.

The evaluation began by an introduction of our work and its context, followed by informing the participants regarding the purpose of the session and their role within it. After presenting the project, participants were introduced to a a future scenario in

Participant	Role	Experience	Education	Age	Gender
1	Operator	5-10 years	High School	30-39	Male
2	Operator	30+ years	High School	50-60	Male
3	Operator	30+ years	High School	55-65	Male
4	Operator	20+ years	High School	40-50	Male

Table 5.3: Demographic information of participants at Holmen from the Evaluation

which they would be asked to act while using the think-aloud method. Participants were also informed of the think-aloud procedure in detail and encouraged to ask questions if anything about the task or the method was unclear.

Once the participants were presented to the prototype, the think-aloud session started. Each participant were asked to think aloud, about their thoughts, feelings or questions when navigating through the interface of the prototype. Some of the participants needed prompts to continue talking about their thoughts, such as 'what are you doing right now?'. When they've completed the think-aloud exercise, the semi-structured interview began, while the participants still had the prototype in front of them. As the nature of the interview questions were semi-structured, follow up questions were asked dependent on the participants previous answers.

To finalize the evaluation session, each participant were asked to fill out their answes on the Van der Laan acceptance scale.

5.4.6.3 Evaluation Analysis

Once the evaluations were completed, we began analyzing the data gathered through think-aloud sessions, interviews, and the Van der Laan acceptance scale. For the Van der Laan scale, each participant's responses were scored on a scale ranging from -2 to +2, organized into two sub-scales: a usability score and a satisfaction score. Five questions contributed to the usability score and four to the satisfaction score. In addition to the original scale, three supplementary questions were included and analyzed separately, these were organized into the third sub-scale: Comprehensibility and trustworthiness.

Each question was first scored individually for every participant on a scale from -2 to +2. Once all individual scores had been recorded, the totals were aggregated across all participants and divided by five for the usability questions, by four for the satisfaction questions, and by three for the comprehensibility and trustworthiness questions, in order to calculate the mean score for each sub-scale. In addition to the mean, the standard deviation was also calculated to provide a representation of the variation and the overall performance across both the usability and satisfaction sub-scales.

The think-aloud and interview notes were analyzed using affinity diagramming. To initiate the analysis, we began by reading through all the notes to familiarize ourselves with the data. Quotes and other noteworthy findings were then individually placed on sticky notes in FigJam. From there, we began organizing the notes into

clusters based on similarity. After iterating on this process by revisiting and re-reading all the notes, certain entries were moved to more appropriate clusters, while others were double-coded and placed in multiple clusters. Once we were satisfied with the selection and organization of the clusters, each was given a descriptive label reflecting its content.

See the result of the analysis in section 6.3.

5.4.7 Follow-up Literature Review

We received positive feedback from the evaluation of the parameter simulation, which enabled operators to perform parameter simulations to explore how outcomes might change when modifying specific parameters. This capability was not identified during the initial literature review; therefore, we conducted a follow-up review to investigate whether existing research supports this functionality.

This subsequent review revealed the presence of explainable AI techniques such as counterfactual and prefactual explanations, which describe how an outcome could differ under alternative input conditions. Incorporating counterfactual explanations has also been shown to increase user satisfaction and trust in the system [59].

Additionally, dedicated 'what-if' analysis tools have been developed, such as the What-If Tool, which allows users to explore hypothetical scenarios through counterfactual reasoning and observe how variations in input data influence outcomes [60].

These approaches closely resemble our parameter simulation visualization. Consequently, the positive findings from our evaluation are supported by existing literature, reinforcing the value of enabling 'what-if' analysis in such systems.

5.4.8 Final Iteration of Design Guidelines and Prototype

Following the analysis of data gathered during the final evaluation, the prototype received positive feedback across the participants. The responses indicated that the design successfully reflected the established guidelines, with operators expressing both understanding of and confidence in the interface. Additionally, the parameter simulation feature received positive feedback.

This finding is further supported by our follow-up literature review, which indicates that counterfactual explanations can enhance user trust, and that dedicated tools have been developed to facilitate such analyses (section 5.4.7).

Based on these findings, we decided to retain all previously established guidelines used in the prototype. In addition, a new guideline was introduced: *The design should provide what-if analysis.* This decision was motivated by three factors: its emergence in both the initial operator interviews and the design workshop, the positive evaluation feedback it received during prototype testing, and the theoretical grounding provided by the follow-up literature review.

Based on operator feedback, several minor adjustments were also made to the prototype. More detailed explanations of what constitutes a desired outcome were added, and the desired outcome was incorporated as an explicit parameter in the interface. The confidence score was also clarified through more comprehensive contextual explanations. Finally, the parameter simulation feature was renamed *what-if analysis* to better align with established terminology and to more accurately reflect its purpose. The final set of guidelines is presented in Section 6.5 and the final version of the prototype is shown in Section 6.4.

6

Results

This chapter presents the results from each stage of the design process. It begins with the findings from the thematic analysis, followed by the outcomes of the design workshop and the final evaluation. Lastly, the final guidelines, their rationale, and the final prototype are presented.

6.1 Thematic Analysis of Operator Interviews

Through several iterative cycles in which the dataset was repeatedly coded, the themes gradually evolved. Ultimately, four main themes were defined, as seen table 6.1. See Figure 6.1 for a representation of the results of the thematic analysis. In addition to the six interviewed operators (P1-P6), selected insights from the operator observed at Södra Cell are included and referred to as P7.

Theme	**Frequency (n)**
Decision Making Under Uncertainty	34
Strategies to Reduce Uncertainty	33
Transparency Needs for Interpreting Uncertainty	28
Trust and Skepticism Toward Uncertain System Outputs	21

Table 6.1: Frequency of themes identified across interviews and observations

Figure 6.1: Representation of the results from the thematic analysis

The following section presents each theme and sub-theme in more detail, accompanied by supporting quotes from the interview material and related insights.

6.1.1 Decision Making under Uncertainty

Decision Making Under Uncertainty is a theme that describes how operators make decisions when working under uncertain conditions. Such uncertainty means that operators often need to act without having all the necessary information available, which can limit their ability to fully anticipate the outcomes of their decisions.

Decision making is an inherent part of the operator work, however, smaller decisions are often made by operators, but bigger decisions often arise from a close dialogue with colleagues and the shift-leader. Rules, routines, and recommendations exist for the processes in the factory, but they are often static. As a result, operators have considerable autonomy in practice and may make decisions that deviate from established routines and recommendations when the situation requires it.

6.1.1.1 Experience-based Decisions

Operators often rely on their experience to predict the outcomes of their decisions. This experience is built over years of working in the plant, recognizing patterns in the DCS values, and identifying deviations in the process. One common decision operators face is determining the right moment to start bleaching, which shifts the pulp quality to a brighter grade. Although a recipe is available to guide this process, it is static, and operators frequently need to adjust it to ensure the pulp meets quality requirements. Experienced operators can draw on their accumulated knowledge when making these adjustments, but less experienced operators often struggle to connect their decisions to past experiences. Additionally, the operators reported that decisions making depends on the specific situation, for some decisions they might rely more on recipes and routines, while other decisions are made out of experience, one operator said:

(P1) *"I use both recipe and experience. Depends on situation. Usually on experience. Sometimes the recipes are not accurate and then it's based on previous experience".*

Another operator said;

(P7) *"When I look at an fault and try to solve it I use my own experience and look at the history. For instance, if there is an extreme value I try to go back by looking at a trendline to see how the value has been before to see if it is common or not".*

Another operator reported the importance of team expertise and dialogues while making decisions:

(P1) *"Dialogue with colleague and boss for harder decisions, depends on the decisions and situation but it is always a dialog together. We have good experience in our team so decisions are made effectively, but I wouldn't say we have any difficult decisions".*

When solving problems under uncertainty, experienced operators primarily rely on their own knowledge and past experience to identify the most accurate solution while

also keeping a close dialogue with colleagues and shift leaders.

6.1.1.2 Rule-based Decisions

All operators rely on rules and recipes to some extent when making decisions. However, because these recipes are not always fully accurate, operators must often deviate from them. Less experienced operators tend to follow the recipe more strictly, which might indicate that they typically lack the practical experience needed to judge when deviations are necessary and what kinds of adjustments are appropriate. When asked about the use of recipes and rules, one operator said:

(P4) *"I only follow the recipes and routines".*

Another operator said:

(P3) *"When encountering a new problem I usually follow the recipe or discuss with my colleagues with higher experience on their thoughts".*

These responses demonstrate how reliance on recipes functions as a compensatory strategy for operators with limited experience: when they lack the practical knowledge to evaluate the situation independently, they turn to structured guidance, or more experienced colleagues, to reduce uncertainty and ensure they make safe decisions.

6.1.1.3 Comfort Levels Effects Decision Making

Operators making decisions under uncertainty use different approaches depending on the type and depth of experience they have. Their level of comfort with these decisions is also closely linked to their experience. Some operators place strong trust in the collective expertise of the team and can make necessary decisions through discussion. Others contact the shift leader directly when faced with a decision. Several operators also point out that even if they are not the ones making every decision, they still play an important role in the decision-making process, as the shift leader does not always possess the specific expertise required for the issue at hand.

We asked the operators how they generally feel when making decisions under uncertainty, and some of their responses included:

(P4) *"I call the boss when something isn't working".*

(P6) *"I do not have any problems with that, as long as I have a motivation for my decision".*

(P4) *"Since I do not make these decisions I feel confident".*

These examples illustrate how operators emotional responses to uncertainty vary, but also how their strategies, whether relying on authority, discussion, or personal reasoning, reflect their experience level and shape their confidence when making decisions under uncertain conditions.

6.1.2 Strategies to Reduce Uncertainty

The theme 'Strategies to Reduce Uncertainty' covers insights regarding the operators work practices to reduce uncertainty in their work. Several of the more experienced operators explained that they work proactively by continuously monitoring the process and detecting changes early, enabling them to take timely action. In contrast, less experienced operators do not perceive uncertainty as a major part of their work. They believe that the established recipes and routines are well suited to the processes, and therefore they rarely deviate from them.

An important insight from the interviews is the clear difference in working approaches depending on the level of experience. Operators with extensive experience often feel confident in uncertain situations and rely on both their own and the groups collective expertise to make wellgrounded decisions. Operators with less experience often directly turn to the shift leader to make decisions.

6.1.2.1 Reliance on Established Routines

As mentioned previously, less experienced operators rely more heavily on the recipes and routines provided to them. This also serves as a clear strategy for reducing uncertainty: by choosing to follow established recipes and routines, the operators minimize the level of uncertainty they experience. One operator commented:

(P4) *"I think the recipes are working very well but I am not so experienced on that, so I never make the deviations".*

Using recipes is also a way to shift responsibility from the individual to the established system. By following the approved method, operators can justify their decisions and feel secure knowing they acted according to the formal guidelines. One of the operators said:

(P3) *"Since we are not formally the ones making the decision, you feel safe knowing its not your own head on the line".*

These insights show that reliance on established routines not only helps less experienced operators manage uncertainty, but also serves as a way to distribute responsibility, providing a sense of security when making decisions in complex or unfamiliar situations.

6.1.2.2 Preventative Work Practices

Several operators engage in preventative tasks to ensure that production runs smoothly and to minimize the risk of failing to meet the pulp quality requirements. The more experienced operators know which values to monitor, what normal conditions look like, and can easily detect malfunctions or deviations from expected patterns. As one operator explained:

(P2) *"I always try to look into the future to see the potential consequences of the decisions I make".*

Less experienced operators may find it difficult to work preventative, as they do not always know the normal operating values and therefore struggle to detect minor issues before they escalate. As one of them explained:

(P7) *"I try to remember values from before. As I am newer it is hard to remember everything, It is easier if you have more experience i guess".*

This illustrates how experienced operators reduce uncertainty by consistently maintaining a forward-looking perspective. By anticipating potential outcomes and identifying early warning signs, they can intervene proactively rather than re-actively. This approach not only decreases the likelihood of disturbances but also strengthens their confidence in handling complex situations, as they are continuously working to stay ahead of potential issues rather than merely responding to them. Additionally, not as experienced operators might face difficulties due to lack of experience.

6.1.3 Transparency Needs for Interpreting Uncertainty

The theme 'transparency needs for interpreting uncertainty' describes the operators requirements for understanding and making sense of uncertainty in AI-based predictive systems. It highlights the specific elements of the predictions that operators must be informed about for the system to be meaningful, trustworthy, and practically useful in their daily decision-making. These needs go beyond simply presenting a numerical prediction; operators require insight into how the prediction was generated, what data it is based on, and how confident the system is in its output. In other words, transparency becomes a prerequisite for operators to evaluate whether they should act on the prediction, question it, or combine it with their own experience. The following sections will present the subthemes to this, and present the operators attitudes and thought on predictive systems and their transparancy.

6.1.3.1 Desire for Explanation of Predictions

For a predictive system to be accurate, trustworthy, and truly valuable to the operators, there is a strong need for the system to provide clear explanations of its predictions. Operators must not only receive the output itself but also understand how the system arrived at that output and what the prediction actually means in the context of their process.

This need for explanation is closely tied to the operators decision making responsibilities. Since operators are accountable for the consequences of their actions, they require insight into the underlying factors influencing a prediction, for example, which input variables were most important, whether the data used was stable, and how confident the system is in its result. Without this form of transparency, predictions risk being perceived as black boxes, making them difficult for operators to trust or integrate into their established work practices. One operator said:

(P6) *"It would be nice with a prediction system, however, I would want to know it takes the right parameters into account. To see if it resonates similarly to how I think".*

Yet another operator described this desire for visualizing this:

(P1) *"To be able to trust the AI I would want to see how it resonates in the same way I do when trying to solve a problem. For example if this combination of parameters has been before, how many time have we had to make changes?".*

This reflects a need for alignment between the operators own mental model of the process and the model used by the AI. For experienced operators, visibility into the parameters becomes a way to judge whether the AI mirrors the heuristic strategies they have built over years of work.

The less experienced operators also emphasized the importance of understanding which parameters the system considers and how it works. One operator stated:

(P1) *"I think it would be interesting to see what happens to the prediction when the raw material or total pulp volume changes".*

This highlights that parameter transparency does not only serve as a trustbuilding mechanism but also as a learning and support function. For operators with limited experience, seeing which variables the AI relies on could help them understand what matters most in the process, guide their attention during monitoring, and gradually develop their own intuition for the system.

6.1.3.2 Need for Certainty and Confidence Indicators

Another crucial aspect of a predictive system is that its level of certainty or confidence is communicated clearly. Several operators emphasized that without explicit information about how confident the system is in its predictions, it would be impossible to use those predictions as a reliable basis for decision-making. Confidence levels help operators judge whether a prediction should be acted upon immediately, treated with caution, or combined with additional process knowledge. Without this transparency, the predictions risk being perceived as unreliable or incomplete, limiting their practical usefulness in production settings. As some of the operators stated:

(P7) *"I would want to see the likelihood of the prediction is leaning in the right way".*

(P1) *"I would want to see what type of correct it is about, and something that shows the accuracy of the prediction".*

(P2) *"It would be interesting to see how correct the prediction is".*

(P6) *"It is important to know if it is uncertain, and I would want know the correctness of the prediction, what is base it's prediction upon".*

Together, these statements highlight that clear communication of confidence levels is not merely a desirable feature but a fundamental requirement for operators to judge the reliability of predictions and incorporate them meaningfully into their decision making processes.

6.1.3.3 Validation Through Historical Performance

One interview question addressed whether operators would find it useful to see the historical performance of the predictive model, that is, how accurate the predictions have been in the past. Operators expressed differing views on this topic. Some considered this information essential, while others felt it offered little value. One operator stated:

(P6) *"It is not so interesting to see the history of the predictions and whether they were correct or not, since the conditions always change. It would give a false sense of security to visualize it".*

In contrast, another operator expressed the opposite view:

(P2) *"It would be interesting to see how often the predictive system has been correct in the past. If this wasnt visualized, I would take the predictions with a pinch of salt".*

Furthermore, one operator explained that it would be valuable not only to see the history of the models predictions but also to understand how other operators have reasoned in similar situations. As the operator expressed:

(P7) *"To help guide my decision making, I would like the AI to show me whether it has encountered a similar problem before and present different options for how others have solved it, so I can see what decisions were made in comparable situations, as I am less experienced it would be valuable to see".*

These contrasting perspectives show that there is significant subjectivity regarding whether historical correctness should be displayed. While individual preferences vary, the majority of operators indicated that visualizing the models past accuracy would be beneficial - though they differed in how important they considered this feature and how they would use it in practice.

6.1.4 Trust and Skepticism Toward Uncertain System Outputs

This theme covers the insights regarding the operators attitude and trust towards a predictive AI system integrated into their work. This is an important aspect when evaluating the use of predictive systems into the industry, as trust is essential for effective and accurate use.

6.1.4.1 Trust Influenced by System Usefulness

Several operators reported that their trust in a predictive system would largely depend on how useful it is in practice, specifically, how correct and accurate it is in the tasks it supports. As one operator stated:

(P1) *"For me to trust it, I would say it would need to work in a correct way".*

As noted through the open ended interview during the observations, one operator reported that they had previously used a predictive system but eventually removed

it because they did not trust it, as it failed to deliver accurate results.

(P7) *"A while ago we had a predictive system but we decided to remove it since the predictions were not accurate, and due to that, we couldn't trust it".*

Operators emphasized that the system must not only function correctly but also provide a clear advantage over existing work practices. In their view, it would only be worth using if it demonstrably improved the decision making process, making it smoother, more efficient, or more reliable than relying solely on human judgement. If the system cannot deliver better performance in these areas, operators expressed little motivation to adopt it, regardless of how advanced the technology might be.

6.1.4.2 Trust Shaped by Understandability of the System

Another aspect that contributes to trust in the system is whether the operators understand how it works. The system needs to be transparent so that operators can evaluate whether its actions and predictions are relevant and reasonable. Several operators also mentioned that if such a system were introduced, they would initially be patient with its results. As two of the operators mentioned:

(P2) *"In order for me to trust it I would try to be patient in the beginning, with it being insecure".*

(P1) *"I would say that its about trusting the system, to test it first to assure its working. What your own experience is with it".*

However, over time, they would expect increasingly accurate and reliable predictions in order to justify continued use. This indicates that trust is not only built through initial transparency but must be maintained through consistent performance as operators gain experience with the system.

6.1.5 Additional Insights from Interviews and Observations

In addition to the themes identified through the thematic analysis, several important insights emerged that were not captured within those themes. These included observational notes on the operators daily work practices, as well as insights from the interviews regarding reflections and opinions of the current DCS system.

6.1.5.1 Observed Work Activities

It was observed that during the day shifts, operators mainly conducted broad, high-level monitoring of the section of the process they were responsible for. They received both visual and audible alarms, and these alarms were the primary triggers for active operator intervention. When no alarms were present, operators generally engaged in conversations with each other while maintaining an overall overview of the screens. Additionally, each shift team had developed its own preferred way of working and expressed confidence that their method was the most effective. This highlights that even within the same plant, operational practices may vary between teams.

When an alarm was triggered, operators responded based on the alarm type, which varied in both sound and color depending on its importance. Their first step was to review the screens for the highlighted values associated with the alarm. They then assessed whether the values were unusual or extreme. If uncertain, operators either consulted with colleagues or reviewed historical trend lines to understand typical behavior. It was also noted that, for most alarms observed on that day, no corrective action was required, and operators simply acknowledged and cleared the alarms.

6.1.5.2 Opinions on Current DCS System

In addition, during the interviews the operators were asked about their attitudes and thoughts regarding the current DCS system.

More experienced operators, who had been using the system for a long time, felt that it functioned well and was easy to interpret.

(P1) *"It is easy to interpret the visualizations, but Ive been working there for a long time. Some chance to customize between screens. We can see all the screen values in curve groups, it allows you to see the progress over a long period of time".*

(P6) *"Pretty easy I would say, probably because of my experience".*

In contrast, the less experienced operators explained that while some visualizations were straightforward, it was more difficult for them to remember previous values or identify deviations.

(P4) *Yes, once you understand how it works it's easy, but there is a lot to keep track of.*

Overall, these responses indicate that while experienced operators find the current DCS intuitive and easy to navigate, less experienced operators face greater challenges in interpreting values and identifying deviations, suggesting that usability is closely tied to familiarity and experience.

6.2 Results of Design Workshop

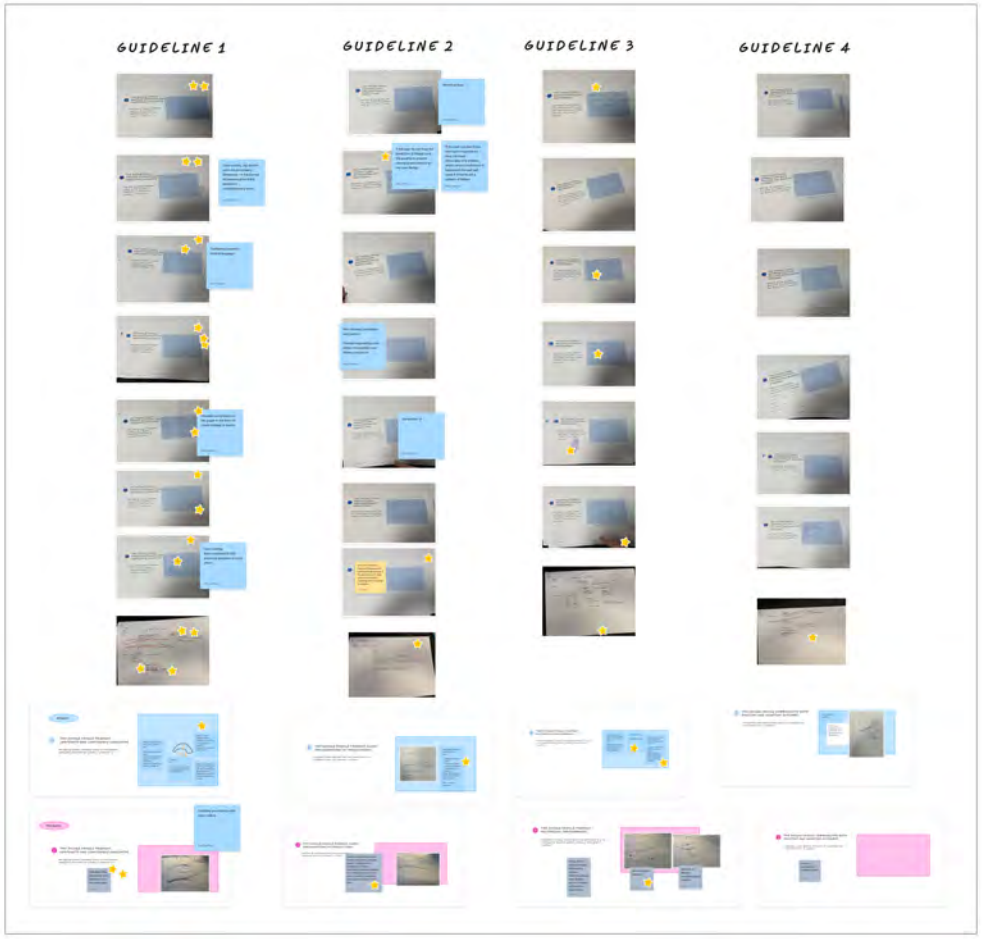


Figure 6.2: Overview of sketches and notes from the workshop

The Design workshop conducted with other students and colleagues at ABB gave us rich insights into how well our guidelines were formulated, as well as how applicable they were to design features of an interface. The workshop provided us with a variety of design approaches for each guideline, and these insights were incorporated as inspiration for the next steps of the prototyping work. Furthermore, the workshop provided us with both engineering and design perspectives. The design students contributed valuable design solutions grounded in the guidelines, while the engineering students focused more on developing technical requirement specifications based on them. See overview of sketches generated from the workshop in figure 6.2. Four guidelines were evaluated, and the insights from each are presented below.

3. The Design Should Provide Certainty and Confidence Indicators

Participants found the guideline easy to understand and helpful for generating design ideas. The resulting design solutions included numerical percentages, color coding, gauges, and widgets.

4. The Design Should Provide Clear Explanations of Predictions

Participants also found this guideline easy to understand and generated multiple ideas for how to design for it, for example in relation to how parameters and model reasoning could be addressed. They suggested that confidence should be communicated using both natural language and numerical data, as well as through explanations of historical data.

Additionally, there were suggestions that it would be interesting to explore how predictions change when certain parameters are adjusted, as well as to include explainability models to support this.

5. The Design Should Provide Historical Performance

This guideline was also considered easy to understand and prompted various ideas for how it could be designed. Participants emphasized that it is important that it does not create a false sense of security. Suggested design approaches included comparison graphs, visualizing data input, and showing contributing factors.

6. The Design Should Communicate Both Positive and Negative Outcomes

Participants found this guideline difficult to understand and challenging to design for. Several participants reported that they struggled to interpret its intent, particularly in terms of how both positive and negative outcomes should be meaningfully represented in a balanced way. As a result, multiple participants did not produce any design concepts for this guideline, as they were unsure how to translate it into concrete design decisions.

6.3 Results of Final Evaluation

This section presents the findings from final evaluation conducted with four participants. The evaluation took place in person together with operators at Holmen Braviken paper mill.

6.3.1 Semi-structured Interview and Think Aloud Results

For the semi-structured interview and the think aloud notes from the evaluation, the data were analyzed using affinity diagramming, resulting in seven thematic clusters. The seven clusters will be presented in the following sections.

6.3.1.1 Trust and Reliability

A consistent finding across all participants was that trust is not granted automatically but must be earned through a demonstrated positive performance of the system. Participants expressed that they would be positive to rely on the system outputs to make decisions, but only once they had observed it performing accurately over a longer period of time. As one participant noted:

(P2) *"personal experience would remain as a central part until the system proves itself".*

One participant raised a distinct concern regarding how the trust is affected by the quality of the underlying data, rather than predictions themselves. This was due to the participants experience of having poorly calibrated sensors and equipment, which he expressed would affect the trust more. This highlights an important distinction between trust in the system and trust in the data foundation.

(P4) *"Our sensors and equipment are not calibrated very often, so there is a lot that is inaccurate. I would find it hard to trust the system because of that- it is just more data. I am skeptical about how things work today, with faulty calibrations and so on."*

Several participants also described that using the system as a form of accountability support, i.e., a way to justify decisions made and provide a documented basis. This suggest that trust in the system is not only an aspect of predictive accuracy, but also about its role in supporting operators own confidence.

6.3.1.2 Historical Accuracy

The feature of being able to see historical accuracy of the system predictions emerged as one of the most valued features among all four participants. Participants expressed a strong preference for being able to see how well predictions aligned with actual outcomes in the past. Most of the participants viewed this function as the primary basis for building trust in the system over time. Additionally, one participant described the feature of historical accuracy as the most helpful feature, noting that it would support adjustments and improve the decision making process.

(P4) *"Historical accuracy is the most helpful feature it would support recipe adjustments"*

Another participant notes that the level of confidence in the historical accuracy would impact decisions:

(P2) *"Low historical confidence would change my decision."*

Several participants also noted that the current process in the paper mill makes it very difficult to review past decisions and their outcomes, which makes the historical accuracy view particularly valuable. This suggest that visualizing prediction history alongside with actual outcomes is a functional requirement for operators.

6.3.1.3 Confidence Interpretation

Most participants interpreted the 42 percent confidence as the probability in which the prediction would match the actual outcome, suggesting that it did communicate the confidence in a clear way. Additionally, participants noted that it was valuable to be able to see the reasons for the confidence number. Furthermore, participants found value in receiving an approximate measure, even though the confidence was low. One participant noted that a 42 percent confidence value, is more reliable than

the manual process they currently use. Overall, the confidence value seemed to be valuable for trusting the prediction, as one participant noted:

(P2) *"Confidence value helps me trust the prediction."*

6.3.1.4 Confidence Needs Explanation and Context

A consistent theme was that confidence values needs to have a contextual framing, otherwise they were seen as meaningless. Participants expressed a need not only to understand the degree of confidence or uncertainty, but also what is actually refers to, for example, whether the prediction falls within an acceptable range relative to the goal brightness. One participant noted:

(P1) *"Confidence number alone doesn't say much without contexts."*

Furthermore, participants also noted that they want explanations for why the system is uncertain, if it is due to limited data, seasonal variation of the raw materials or calibration issues. Without that information the participant expressed that the confidence indicator would be insufficient as a standalone decision support. This suggest that confidence and uncertainty should be communicated alongside with contextual annotations.

6.3.1.5 Parameter Importance

Another theme that emerged was parameter importance. Participants reported that explicitly presenting the parameters used in the prediction increased their trust in the system, as it helped them understand how the model operates. Additionally, participants found parameter simulation valuable for understanding the prediction model and how it behaves. They suggested that parameter simulation increased their trust in the predictions, as it provided insight into how outputs change when inputs are varied. This also supported a better overall understanding of the model.

6.3.1.6 Uncertainty Visualization

One of the main features of the prototype interface was the uncertainty visualization around the predictions solid line. The participants generally responded positively to the visualization, with several participants noting that it felt intuitive and reasonable. The blurred area surrounding the prediction line was interpreted as the representing margin of imprecision, which clearly aligns with the visualizations intended meaning. One participant described it as the following:

(P3) *"I would say that the blue blur around the solid line represent the margin in which the outcome may fall."*

However, one participant noted that it was difficult to interpret the blur around the prediction line, additionally the participant noted that it was valuable to have both the visualization alongside with the confidence percentage which made the interpretation more sufficient.

(P4) *"Do not understand what the colors around the curve mean."*

Another aspect were that participants had a clear preference for seeing both the predicted value and the actual outcome in the same view, to enable a more smooth function of comparing. One participant said:

(P2) *"It's easy to understand but I prefer to have both the graph and the percentage of confidence."*

6.3.1.7 Decision Making and System as a Complement

Several participants expressed a positive attitude to use the system as a tool rather than only relying on the system. One participant, who had no to little experience in the domain noted that they would have to rely on the system, highlighting how operator expertise is shaping reliance behavior:

(P1) *"I don't have so much experience of this, so I would have to rely on the system."*

Additionally, more experienced operators framed reliance as a conditional of the system first proving its accuracy in practice. Furthermore, several participants expressed interest in receiving additional contextual information to support decisions, such as start time per bleach point, target brightness spans and seasonal parameters. While this is not directly related to uncertainty communication, this need reflects a broad expectation of what the system should provide to be a complete decision making tool.

6.3.2 Van der Laan Acceptance Scale

To evaluate the acceptance of the uncertainty visualizations, the Van der Laan acceptance scale was administered to four participants (n=4). The scale measures acceptance across two dimensions, usefulness and satisfaction, each scored from -2 to +2. The original scale consist of nine items, and addition to those, three supplementary items were added to assess comprehensibility and trustworthiness.

The results showed that the uncertainty visualization was rated positively on all three dimensions. The usefulness sub-scaled had a mean score of **M=0.9 (SD = 0.93)**, and the satisfaction sub-scale had a mean score of **M=1.0 (SD = 0.82)**, both indicating a generally positive acceptance of the visualization. The supplementary sub-scale for comprehensibility and trustworthiness had a mean score of **M=0.67 (SD = 1.09)**, suggesting a moderately positive perception, though with a greater variation between participants.

It should be noted that one participant consistently rated the visualization negatively across all dimensions, which contributed to the higher standard deviations observed. Additionally, given the small sample size (n=4), these results should be interpreted with caution and are considered as indications rather than conclusions. See table 6.2.

Sub-scale	Mean	Standard Deviation
Usability	0.9	0.93
Satisfaction	1.0	0.82
Comprehensibility & trustworthiness	0.67	1.09

Table 6.2: Van der Laan acceptance scale summary

6.4 Final Prototype

The final prototype is illustrated through screenshots of the forecast view and the historical accuracy view, including a version of the latter where all tooltips are displayed simultaneously (see Figures 6.3, 6.4, and 6.5).

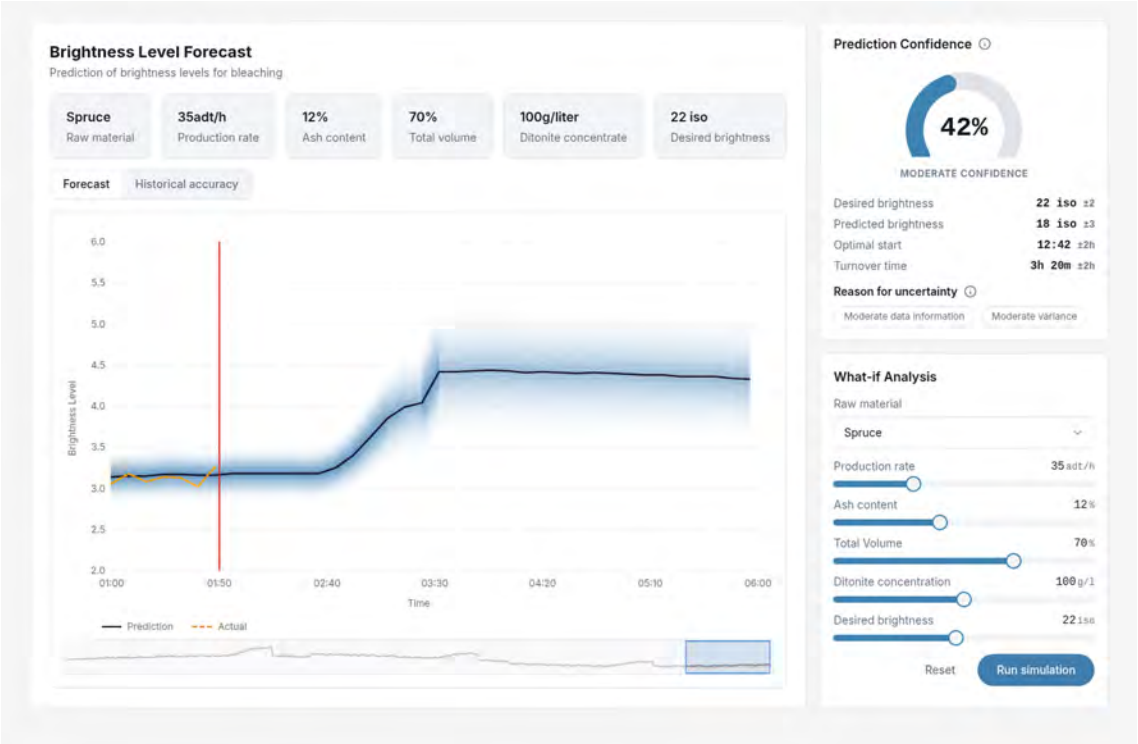


Figure 6.3: Final prototype showing the main interface (forecast view)

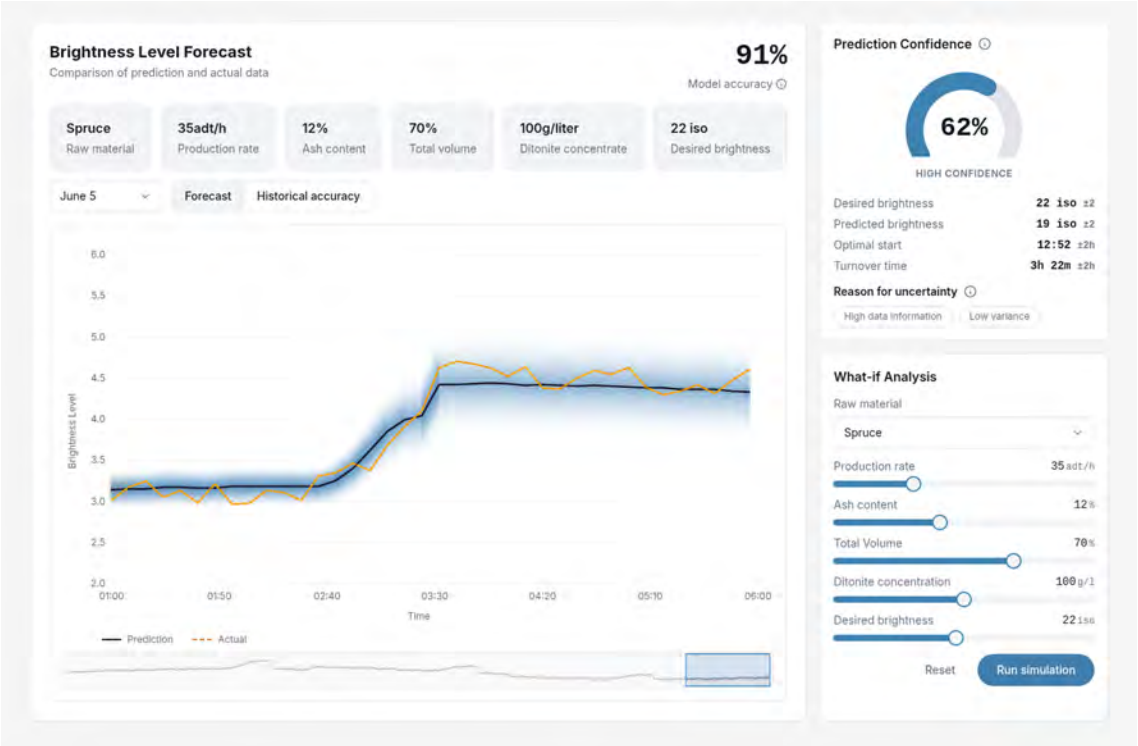


Figure 6.4: Final prototype showing the historical accuracy view

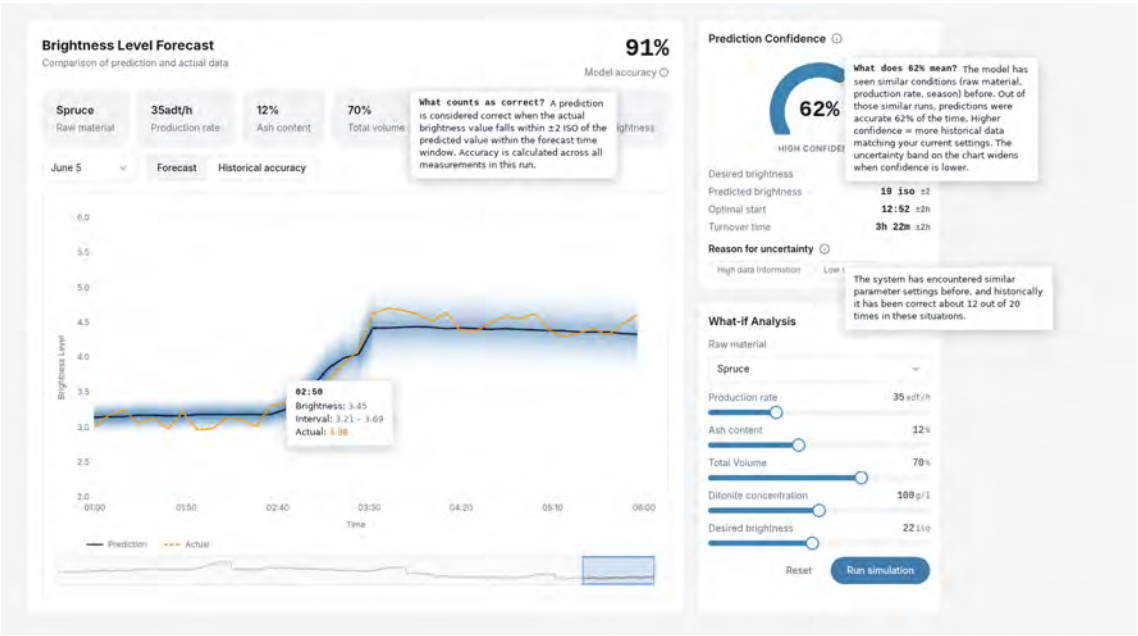


Figure 6.5: Final prototype showing the historical accuracy view with all tooltips displayed simultaneously

6.5 Final Design Guidelines

The design guidelines presented in this section represent the final, validated outcome of an iterative process spanning four sources of evidence: a literature review of uncertainty communication, semi-structured interviews with paper mill operators, a design workshop with designers and engineers at ABB, and a final evaluation with paper mill operators. An initial set of guidelines was formulated during the Define phase (Section 5.3.2) and used as a foundation for prototyping and the workshop. Following the workshop and final evaluation, the guidelines were reviewed and refined. One guideline: *the design should provide positive and negative outcomes* were removed after the design workshop. And one guideline: *the design should provide what-if analysis* was added in the final iteration, having emerged in both the initial interviews and the workshop before receiving strong support in the evaluation. In addition the new guideline was also supported by an ad-hoc literature review.

Nr.	Guideline	Description
1	The design should be based on users context, experience, needs, and tasks	People interpret uncertainty differently based on their background and expertise. Experienced operators use established mental models, while less experienced ones rely on guidance and routines. Tailoring visualizations to these different needs ensures uncertainty is communicated meaningfully and supports better decisions.
2	The design should aim to reduce cognitive load and clutter	Human short-term memory is limited, making information overload a risk in complex decision environments. Beyond just limiting information, design should provide clear reference points and contextual cues.
3	The design should provide certainty and confidence indicators	Clear confidence indicators help users understand prediction reliability and make informed decisions rather than blindly trusting output. Operators specifically emphasized this as fundamental for assessing reliability and integrating predictions into their decision-making.
4	The design should provide clear explanations of predictions	Users need to understand how the system reasons, main parameters it uses, and historical accuracy. This transparency enables operators to verify the system's logic aligns with their own understanding, building trust and supporting well-grounded decisions.
5	The design should provide historical performance	Operators expressed that the key to trust the predictions is by presenting history with context about the conditions under which predictions were made, so operators can judge relevance. Showing how similar situations were handled previously also builds confidence.
6	The design should provide what-if analysis	More specifically, the ability to interactively adjust input data and observe the effect on the predicted outcome. Allowing users to simulate different scenarios supports a more active and exploratory form of decision-making, enabling operators to interrogate the system's reasoning rather than passively receiving its output.

Table 6.3: Final design guidelines for uncertainty communication in predictive systems

6.5.1 Rationale of Design Guidelines

1. **The design should be based on users context, experience, needs, and tasks**.

   The literature on uncertainty communication establishes that people interpret uncertainty differently depending on their background, expertise, and task demands. Adapting visualizations to users goals, constraints, and mental models has been shown to reduce confusion and support more informed decisions [16], [17], [20], [23]. This principle was consistently reinforced across all phases of this study. In the initial interviews, operators demonstrated markedly different working practices depending on their level of experience: experienced operators applied preventative strategies, drew on established mental models, and exercised independent judgment, while less experienced operators relied more heavily on static recipes, peer guidance, and shift-leader authority to manage uncertainty (Section 6.1.1).

   In the final evaluation, this guideline was confirmed through the observed differences in how operators interacted with the prototype. Participants with less experience indicated greater reliance on the system, whereas more experienced operators framed their willingness to use it as contingent on first observing its performance over time (Section 6.3.1.7). Together, these findings show that effective design in this context should accommodate varying expertise levels and work practices to ensure that uncertainty is communicated in a way that is both meaningful and actionable for each user.

2. **The design should aim to reduce cognitive load and clutter**

   Research on uncertainty visualization emphasizes that simplifying visual representations, removing irrelevant detail, and optimizing layout are essential for supporting comprehension [20]. This is underpinned by cognitive theory showing that human short-term memory is limited, making users particularly vulnerable to information overload in complex, high-stakes decision environments [9].

   In the initial interviews, operators did not report feeling overwhelmed by information volume per se; but rather struggled to remember previous values and lacked clear reference points for determining what constituted a normal or desirable value (Section 6.1.5.2). The final evaluation further reinforced this: participants expressed a need for confidence values to be accompanied by contextual framing; for example, whether a given prediction falls within an acceptable range relative to the target brightness, rather than being presented as standalone numerical outputs (Section 6.3.1.3). This suggests that cognitive burden in this context arises not only from excess information, but also from the absence of contextual anchors that allow operators to interpret process values without relying entirely on memory. This insight directly informed several prototype decisions, including the addition of a desired outcome indicator and more explicit explanations of the confidence score following evaluation feedback (Section 5.4.8). Reducing cognitive load, therefore, is not only

about limiting content but about designing interfaces that provide meaningful context.

3. **The design should provide certainty and confidence indicators**

   The literature establishes that including certainty and confidence indicators in predictive system outputs helps users interpret the reliability of predictions and make more informed decisions. Clear confidence information has been shown to promote appropriate trust, encouraging users to engage analytically with AI-supported recommendations rather than accepting them uncritically [2], [5], [14], [18]. Another important aspect is that many people struggle to interpret numerical information, which makes it difficult for them to fully understand the uncertainty that such data conveys. Research has shown that numeracy levels significantly influence how individuals evaluate and respond to risk-related information [15], [16].

   The initial interviews confirmed this at the level of operator need: operators emphasized that clear communication of confidence levels is a fundamental prerequisite for assessing the reliability of predictions and integrating them meaningfully into decision-making (Section 6.1.3.2).

   The design workshop produced multiple design solutions for representing confidence, including numerical percentages, gauges, and color-coded indicators, reflecting the participants' recognition that this guideline required concrete and varied expression in the interface (Section 6.2).

   The final evaluation provided the most direct validation of this guideline. Participants generally interpreted the confidence value in the prototype correctly, understanding it as the probability that the prediction would match the actual outcome (Section 6.3.1.3).

4. **The design should provide clear explanations of predictions**

   The literature recommends that the design should provide statistical information and model information in order to help the users make informed decisions. Statistical information helps users understand the likelihood associated with different outcomes and correctly interpret uncertainty intervals. In addition, information about the model, such as relevant input parameters, their influence, and the systems historical accuracy, enables users to evaluate the reliability [20].

   This guideline was also supported by insights from the interviews. Several operators emphasized that, in order to make wellgrounded decisions and develop trust in a predictive system, it would be essential to understand how the system reasons and which types of information it relies on. They expressed a need to see the key parameters behind a prediction and to verify that the systems logic aligns with their own understanding of the process, see further details in section 6.1.3.1.

   The design workshop translated this guideline into concrete interface proposals, with both design and engineering participants generating solutions for how pa-

rameters and model reasoning could be surfaced within the interface (Section 6.2). The prototype addressed this through input parameter cards displayed in the main view, alongside contextual info icons that provided on-demand explanations of the confidence percentage and the sources of uncertainty.

The final evaluation confirmed this approach: participants responded positively to being able to see the parameters behind the prediction (Section 6.3.1.5). It was also found that participants wanted even more comprehensive explanations of the confidence score (Section 6.3.1.4).

5. **The design should provide historical performance**

   As mentioned in previous design guideline, there is a need to present historical accuracy and performance in prediction [20]. Our initial interview findings show that operators have differing attitudes toward viewing historical prediction accuracy. Some operators considered this information valuable and even expected it to be included, whereas others argued that changing process conditions make such history less relevant and that showing past performance might create a false sense of safety, see more details in section 6.1.3.3. Because of these different perspectives, it is essential that any visualization of historical performance does not merely present raw accuracy, but also communicates the conditions and parameters under which previous predictions were made. Providing this contextual framing can help operators judge whether the historical cases are comparable to the current situation, ensuring that the information supports rather than misleads their decisionmaking. See more details in section 6.1.3.3.

   The design workshop generated multiple design suggestions on how historical performance could be integrated into the interface without creating a false sense of security, generating proposals that emphasized contextual framing alongside accuracy metrics (Section 6.2). This directly influenced the prototype design, in which historical accuracy was placed in a dedicated secondary view and supplemented with an indicator allowing operators to identify cases where prediction confidence and actual outcomes diverged.

   In the final evaluation, historical accuracy emerged as one of the most valued features across all four participants. Most participants viewed it as the primary mechanism for building trust in the system over time. (Section 6.3.1.2).

   These findings confirm that the divergent perspectives identified in the initial interviews do not argue against including historical performance, but rather argue for how it must be presented: always accompanied by contextual information about the conditions under which previous predictions were made, so that operators can judge the relevance of historical cases to their current situation.

6. **The design should provide what-if analysis**

   In the initial interviews, several operators expressed a desire to explore hypothetical scenarios - for example, to understand how a change in raw material

composition or total pulp volume would affect the brightness forecast (Section 6.1.3.1). This interest was not captured by any of the existing guidelines at the time, but was noted as a recurring theme. During the design workshop, participants similarly raised the idea of interactive parameter adjustment as a meaningful way to make the system's reasoning tangible and to support exploratory decision-making. Although it was not a formally evaluated guideline in the workshop, the interest it generated among participants reinforced its potential value (Section 6.2.

To further investigate this, the parameter simulation feature was incorporated into the final prototype to gather evaluative feedback on its usefulness. In the evaluation, participants responded positively to the ability to adjust input parameters and directly observe the resulting changes in forecast outputs (Section 6.3.1.5). A follow-up literature review supports these findings, indicating that counterfactual explanations can increase trust in a system [59], and that dedicated tools exist to facilitate exploration of hypothetical scenarios through counterfactual reasoning [60].

7

Discussion

This chapter synthesizes the findings of the study and reflects on their broader meaning in relation to the research questions, the methodology, and the wider field of human-AI interaction. The chapter is structured as follows. First, the three research questions are addressed in light of the findings from interviews, the design workshop, and the final evaluation. This is followed by methodological reflections on the strengths and limitations of the chosen approaches. The chapter then discusses the broader implications of the work for AI design in industrial settings, before addressing the ethical considerations that arise when introducing AI-based decision support into safety-critical environments.

7.1 RQ1 - How do operators interpret and reason about uncertainty in their current work?

Through the initial interviews and observations, we investigated how operators in their current work handles, reason, and interpret uncertainty. What we can tell of this investigations is that operators interpret uncertainty differently, and this variation is closely tied to their level of experience. More experienced operators have, over time, developed a robust foundation of mental models that guide them through uncertain situations, allowing them to draw on past experiences where conditions were comparable. Less experienced operators, while not denying the existence of uncertainty, tend to perceive it as a less prominent feature of their work, managing it primarily by adhering strictly to established routines and procedures.

The operators' current work practice is highly based on routines and recipes, and less experienced operators tend to rely more on these, than more experienced operators. This finding closely aligns with Rasmussen's SRK framework [12], where experienced operators increasingly operate at a skill-based level guided by internalized mental models, while less experienced operators rely on rule-based behavior governed by established procedures.

Another important aspect is that uncertain situations are typically handled collectively, operators maintain close dialog with shift leaders and colleagues, drawing on the shared expertise and experience of the team.

7.2 RQ2 - How does uncertainty communication influence trust, reliance, and decision-making?

The findings from both the initial interview and the final evaluation suggest that uncertainty communication has a significant influence on operators' trust, reliance, and decision making. Trust in such a system was not found to be a single, static quality but rather something that is earned through continued use, where successful interactions, and successful predictions is important. If a system clearly demonstrates a positive and efficient prediction output, then operators would feel safe using such a system. However, beyond the output quality of the AI model, trust was also closely tied to how well operators could understand the systems reasoning. Rather than simply wanting a single prediction, operators wanted insight into the underlying logic, assumptions and basis for the output in order to critically evaluate it. While this broadly aligns with Prabhudesai et al. [2], who found that communicating uncertainty encourages more analytical engagement with AI outputs, our findings points to a strong dynamic in this industrial context. Uncertainty information alone was insufficient, operators required access to the systems underlying logic and reasoning as a precondition for integrating predictions into their decision making at all. By comparing the system's reasoning against their own mental models, operators could assess whether the system's conclusions felt reasonable and aligned with their own understanding of the situation. This need was particularly prominent among the more experienced operators, who had developed strong intuitions about the bleaching process over years of practice. For these operators, a system that does not show its reasoning visible, risked being perceived as a 'black box', and would give a false sense of security. This also aligns with existing research claiming that transparency is a prerequisite for appropriate reliance in human AI interaction [5], [14]. This need for transparency extended to less experienced operators as well, who equally valued insight into the system's reasoning. However, due to their more limited experience, they were less confident in their ability to critically evaluate it - lacking the deep operational knowledge needed to compare the system's logic against a well-formed mental model of their own.

Confidence indicators was found to be a fundamental requirement rather than an optional feature. Across both interviews and evaluation, operators consistently expressed that without explicit information about how certain or confident the system was in its predictions, they would be unable to integrate those predictions into their own decision making. The confidence level functions as a signal that allow the operators to calibrate their decision, determining whether to act immediately, treat the prediction with caution or fully rely on their own experience. This finding is consistent with Cao et al. [19], who demonstrated in a medical context that presenting calibrated model uncertainty helps users adjust their reliance more effectively. Additionally, their study also found that calibrated uncertainty alone was insufficient, users needed the uncertainty to be framed in an interpretable format to make meaningful use of it. Our findings echo this into the industrial context, operators did not only need the confidence indicator, but required it to be contextualized within their operational reality.

The display of historical prediction accuracy was viewed differently across operators. For some, it was an essential feature that helped build confidence in the system. Others, however, warned that it could create a false sense of security, pointing to how continuously shifting parameter values can have significant effects on outcomes, making past accuracy potentially misleading when conditions differ. This concern is consistent with Zimmermann [4], who highlights that humans are prone to overconfidence when interpreting model outputs, risking a false sense of security regarding their reliability. That risk was raised by operators themselves, which lends empirical weight to what has previously been a theoretical concern. To address this concern, we decided that any historical accuracy shown should only reflect predictions made under the same parameter conditions as the current one, ensuring a more meaningful and honest basis for comparison. However, during the evaluation, most of the operators were positive to the function, and it was found to be one of the most important features of the prototype tested in the evaluation.

Taken together, these findings suggest that beyond having an AI model that works accurately, effective uncertainty communication is a prerequisite for adequate reliance. When uncertainty is clearly communicated, together with historical accuracy, operators holds a better position to decide when to trust the system, when to handle it with caution and when to fully base their decision on their own expertise.

7.3 RQ3 - What design guidelines support trust and understanding of AI uncertainty?

The design guidelines developed in this work emerged iteratively through a user centered design process, grounded in insights from literature, operator interviews, the design workshop at ABB, and the final evaluation with the operators. Rather than prescribing a visual solution, the guidelines aims to reflect as a set of principles that together support operators in the process industry in forming trust, and helping them making informed decisions when interacting with AI systems.

A fundamental principle that runs through most of the guidelines is the need for visible uncertainty communication. Operators need to have all information about the prediction visible in the interface, such as the uncertainty level, what it means, what the system base its prediction upon and the historical accuracy of similar predictions. This finding is consistent with the findings in the literature, which emphasizes that uncertainty should be treated as a primary information layer rather than a detail [5]. Additionally, the confidence level needs to be contextualized in terms that are meaningful within the operational environment. A percentage value of confidence alone doesn't say much for the operators, they often require additional explanations to interpret what a given confidence level actually meant in practice. This finding is directly consistent with Bhatt et al [16], who show that people may struggle to interpret numerical risk data and recommended explaining uncertainty in words and as meaningful ranges. This suggest that numerical indicators of uncertainty need to be accompanied by qualitative framing or contextual cues that translate abstract probability into actionable operational meaning. Additionally, Padilla et

al. [23] emphasize that there is no single visualization technique that works for all uncertainty cases and that poor design choices can add confusion. In our design approach, we offered multiple layers of uncertainty information, which is a deliberate response to this principle.

Transparency about the systems reasoning also emerged as a central design guideline. Operators, especially experienced ones, expressed a strong desire to understand which parameters the system considered when generating a prediction, and whether that reasoning aligned with their own understanding of the process. Designing for this need does not necessarily require exposing the full technical details of the model, but rather providing enough insight into the systems logic so that operators can assess its relevance and validity in a given situation. For less experienced operators, the transparency serves an additional function, it can act as a learning support, helping them understand which process variables are most significant and eventually, this might help them develop their own intuition.

The guidelines also address the role of historical performance in supporting trust calibration. Operators held diverging opinions through the interviews, suggesting that the features value is highly subjective. However, keeping it as a part of the interface was showed to be valuable through the evaluation, without making it the primary focus of the interface. Additionally, the historical accuracy needs to be framed carefully, with clear indications that past performance does not guarantee future accuracy.

One guideline that showed to be particularly interesting, was the 'what if' feature that allow users to simulate predictions by adjusting parameter values of their own choice. Its inclusion was driven by a convergence of evidence across multiple phases of the study: operators expressed a desire for it in the initial interviews, designers and engineers independently identified it as valuable during the workshop, and it received positive feedback in the final evaluation. The subsequent literature review then confirmed that this need is supported by established XAI concepts such as counterfactual and pre-factual explanations [59]. The operators noted that the feature should serve a dual purpose: evaluating model behaviour and supporting less experienced operators in understanding how parameter values influence outcomes. This aligns closely with Wexler et al. [60], who position 'what-if' analysis as a mechanism for users to construct hypothetical scenarios and observe the effect of varying input data. This trajectory, from operator insight to design validation to theoretical grounding suggests that the what-if analysis guideline captures something that current uncertainty communication research has not yet fully addressed in industrial contexts. Rather than simply presenting a static prediction, enabling operators to actively interrogate the model by adjusting input parameters transforms the system from a passive output device into an exploratory tool. This supports a more reflective and informed form of decision-making, and aligns with the broader principle that AI systems in high-stakes environments should complement operator expertise rather than replace it.

To summarize, the guidelines reflect the principle that the system should position itself as a complement to operator expertise rather than a replacement for it. The

interface should make clear that the prediction is only a prediction, and that the operators own judgment remains central to the decision making process

7.4 Methodological Reflections

The following sections reflect on our methodology, evaluating how well our chosen methods contributed to the design process and what we would have done differently. While most methods proved genuinely valuable, some fell short of their intended purpose.

7.4.1 Literature Review

The literature review conducted in the early stages of this work was fundamental in establishing a clear picture of the existing research landscape - what prior studies have shown, what approaches have already been explored, and crucially, where gaps in knowledge remain. Beyond mapping the research area, the literature review was equally important in grounding the design guidelines in existing research on uncertainty communication. It also provided valuable insight into how uncertainty communication has been approached outside of the specific industrial context studied here. While this broader perspective offered a solid foundation for understanding what has proven effective in other settings, much of the available research stems from different domains, which makes direct translation to the process industry more challenging and needs careful consideration.

7.4.2 Initial Interviews

Semi-structured interviews were conducted in the initial phase of the design process to gain insight into operators' current work practices and to understand how they reason and act in uncertain situations. Through these interviews, rich qualitative data were gathered that deepened our understanding of the operational context. The majority of participants were engaged and provided detailed, elaborate responses. However, some operators primarily those with less experience in their role, found certain questions more difficult to answer, which could largely be attributed to their more limited operational experience.

If we had more time, and more participants, we would have preferred to conduct a pilot interview, to further evaluate the quality of the questions and to see if they were formulated in a way that made sense to the operators. However, the choice of keeping the questions semi-structured allowed us to adjust our formulations and questions for each participant, allowing us to ask follow-up questions if something was unclear or if we wanted more information.

7.4.3 Design Workshop

The design workshop was highly valuable in the development phase of the process. Given that the primary deliverable of this work is the guidelines, we considered it to

be important to evaluate these together with other students in the fields of UX and engineering, as well as UX researchers at ABB. It was particularly valuable with the engineering perspective, as their way of thinking differed significantly from that of the designers. They tended to frame the guidelines in terms of concrete system constraints and concepts, such as thresholds, parameters, and measurable criteria, which complemented the more solution oriented and exploratory approach of the design students.

For this workshop, we used 'how might we' questions, based on four of our six guidelines. Two of the guidelines was excluded from this exercise since they were too broad, and served as a more general approach to the design, rather than design features. However, the 'how might we' exercise proved well suited for framing the design guidelines in an open and exploratory manner, and generated a broad spectrum of ideas, while also providing us with external perspectives that enriched our own assumptions.

However, the 'How might we' questions may have steered participants in a certain directions without us being aware of it, questions are never fully neutral and could without intention limit the creative space.

7.4.4 Prototyping

The base of the prototype was designed in Figma, where all components, features and visual elements were crafted based on each of our formulated guidelines. To make the prototype interactive and testable in the evaluation session, Lovable was used. Lovable is an AI powered development tool[58], and we used it to translate the static Figma design into a interactive interface.

Using AI tools in the prototyping phase raises valid questions about the design. One concern is that AI generated outputs may introduce unwanted design decisions that deviate from the original content. However, since all visual design decisions had already been finalized in Figma, the risk of AI influencing the design itself was limited. Lovable was used strictly to implement interactivity, and not to make any design decisions.

Additionally, from a practical standpoint, using Lovable increased the efficiency of the prototyping process, allowing us to fully focus on the evaluation rather than technical implementation.

7.4.5 Final Evaluation

The final evaluation took place in person with operators at Holmen Braviken paper mill, with the goal of assessing whether our design guidelines aligned with operator needs, tested through a high-fidelity prototype. The methods used were think-aloud, semi-structured interviews, and the Van der Laan acceptance scale.

The think-aloud method was chosen to capture real-time reactions: questions that arose naturally, moments of confusion, and positive impressions of the interface. Participants were given a clear task description before the session began, but most

struggled to verbalize their thoughts and remained largely silent throughout, requiring repeated prompting to continue. As a result, the method yielded less feedback and fewer insights than anticipated.

In hindsight, running a brief trial exercise beforehand would have been beneficial, giving participants a chance to grow comfortable with the method before engaging with the actual prototype.

Additionally, we chose to have the Van der Laan acceptance scale as a method for our evaluation to gather some quantitative data, but due to our small sample size (N=4), we are unable to draw any conclusions based on those results. However, we still believe it was valuable to include that method since it gave us some indications of the participants attitudes towards usability, satisfaction, comprehensibility and trustworthiness.

Furthermore, several participants appeared confused by the concept of AI itself, which led them to ask questions about the underlying technology rather than engaging with the visualizations that are the core focus of this work. This suggests that AI as a concept may still feel too abstract or unfamiliar to operators in this context, making it difficult for them to relate to and evaluate the design on its own merits. This conceptual barrier likely influenced their responses and may have skewed the results, as participants were not always assessing the interface itself but rather grappling with the idea of AI-driven decision support as a whole.

A notable limitation of the evaluation is that the majority of participants were highly experienced operators, with only one participant having limited experience in the role. As a result, the interface was not fully tested across different levels of user experience, and it is therefore difficult to draw conclusions about how well the design supports operators at varying stages of their professional development.

7.5 Implications and broader relevance

Through an iterative design process, this work has focused on exploring how uncertainty can be effectively communicated in AI interfaces for users within the process industry. Conducted in collaboration with ABB Corporate Research Center, the findings carry meaningful implications for the future integration of AI in industrial settings, and it is therefore essential to discuss how these results can inform and shape that trajectory.

One of the main deliverables of this thesis is the design guidelines, which have a solid foundation in existing research and have been tested through interviews, workshops, and evaluations together with UX practitioners, engineers, and operators within the process industry. The guidelines aim to serve as a support tool when designing AI system interfaces in the process industry, with a particular focus on uncertainty visualization.

7.6 Ethical Considerations

The integration of AI systems in the process industry raises several ethical considerations that are crucial to address, especially in safety-critical environments where AI-supported decisions can have significant consequences.

This section highlights three key ethical aspects: AI Accountability addresses challenges around accountability when many actors are involved in the development and use of AI systems. Non-transparent and Unexplainable AI discusses the black box problem and the EU AI Acts transparency requirements. Finally, Operator Autonomy emphasizes the importance of preserving operator autonomy and the risks of over-reliance on AI systems.

7.6.1 AI Accountability

A central challenge is the distributed nature of AI development and implementation. Many actors are involved: from those who develop the models, to those who integrate them into operational systems, and further to users who rely on the models in their daily work. This diversity of actors makes it difficult to investigate the distribution of responsibility when something goes wrong.

The issue becomes particularly relevant in the context of uncertainty communication. If an AI system fails to clearly communicate its uncertainty to the operator, the operator risks making decisions without sufficient evidence. Therefore, it is important to design clear guidelines for how AI systems should transparently communicate their limitations to the user. The guidelines developed in this work directly address this. By establishing concrete principles for how uncertainty should be communicated in AI based prediction systems, they provide a foundation for assigning clearer responsibility for designers and developers. The guidelines, grounded in both existing literature and empirical user studies with domain experts, they aim to ensure that uncertainty is not an afterthought in system development, but a primary design consideration, one that can be documented, evaluated and held to account.

Within the framework of Human-Centered Explainable AI (HCXAI), designers and developers have an ethical obligation to ensure that uncertainty is communicated in an intuitive, understandable, and actionable way for the operators using models [61].

7.6.2 Non-transparent and Unexplainable AI

Many machine learning models make predictions by analyzing complex patterns that are difficult for humans to fully understand. As a result, the reasons behind decisions generated by these algorithms can be unclear to the people they affect, also referred to as the 'black box' problem.

Moreover, the EU AI Act, in its Articles 13 and 14, explicitly requires transparency and human review for AI systems classified as high-risk. Article 13 states that these systems must be designed with sufficient transparency to enable users to adequately

interpret and understand the systems results. The aim is to ensure that users can use AI-generated outputs in an informed and responsible manner, which is particularly critical in contexts where incorrect decisions can have serious consequences.

Article 14 complements these requirements by stipulating that high-risk AI systems must allow for effective human review throughout their lifetime. This means that users must have access to tools and methods that support them in correctly interpreting and assessing the systems recommendations and decisions. Taken together, these requirements are designed to directly address the black box problem and ensure that AI systems do not make decisions that users lack the ability to review, understand or question [62].

In light of these regulations, this work aims to actively comply with the requirements of the EU AI Act on AI systems by developing intuitive and explanatory communication methods to increase transparency around uncertainty in AI predictions. By designing visualizations and interaction designs that support operators' understanding of AI outputs, the work contributes to meeting the requirements for both transparency (Article 13) and human control (Article 14).

7.6.3 Operator Autonomy

A central ethical aspect of this work is to ensure that operator autonomy and expertise are preserved when AI systems are introduced into the process industry. As highlighted in section 3.4.1, AI should serve as a support for decision-making, not as a replacement for human expertise. One risk with AI systems is that operators may develop excessive trust in the system's recommendations, which can undermine their ability to make independent judgments [3] [4]. The goal with designing AI interfaces should therefore not be to maximize user trust in the system, but rather to empower the user to know when to trust it and when to treat its outputs with caution.

This work therefore aimed to ethically emphasize the irreplaceable function of humans in the process industry. By developing communication methods that clearly convey uncertainty in AI predictions, the project aims to promote a balanced collaboration between operators and AI models where operators retain control and ultimate responsibility for decisions.

7.6.4 The Use of AI in This Work

The artificial intelligence tool ChatGPT [63] was used to support language refinement and improve the grammatical accuracy of this report. Additionally, the co-creation tool Lovable [58] was used to transform our designs into an interactive prototype. Claude [57] was also used to assist in generating prompts for Lovable. Using AI in this way comes with risks, most notably that AI-generated or refined text may alter the intended meaning or introduce inaccuracies. To address this, we carefully reviewed all AI-assisted outputs before incorporating them, ensuring that the final text accurately reflected our own analysis and conclusions.

7.7 Limitations and Future Work

The initial interviews were conducted with six operators, and the subsequent evaluation with four operators, all employed at the same paper mill. As such, the findings of this study should be interpreted as indicative rather than conclusive. Future research would benefit from a larger sample size, or alternatively, from including operators across multiple mills within the process industry.

A further limitation concerns the contextual scope of this study. This research primarily focuses on a single process, the TMP bleaching process at Holmen Braviken paper mill. Consequently, the guidelines produced through this study may not be directly applicable to other facilities, and it remains unclear to what extent the findings can be generalized to the broader process industry. Future studies may therefore consider expanding the scope to include multiple mills.

The final evaluation was conducted using a high fidelity prototype, developed on the basis of the guidelines established in the earlier stages of this work. Although the prototype was interactive and built upon fictitious data, there remains a risk that operators' attitudes and reactions may differ from those observed in a real world setting, where actual risk are present. Due to this, it is important that prototypes of this kind are continuously tested across multiple studies, incorporating a variety of risk scenarios. A further potential limitation is that the participating operators had no prior experience with predictive systems, which makes it difficult to measure trust in a meaningful and reliable way, as the reactions observed are likely to reflect hypothetical attitudes rather than genuine responses grounded in real-world experience. In addition, the prototype was tested without the operators having any prior exposure to it. This is relevant because, in practice, if a predictive AI system were to be integrated into the process industry, operators would likely undergo structured training before using it. As a result, the conditions under which the prototype was evaluated differ from those of a real deployment, where operators would be expected to have familiarized themselves with the system beforehand. This gap between the test conditions and real-world implementation should be considered when interpreting the findings.

Based on these limitations, several directions for future work can be identified. An important next step could be to conduct studies in real-world settings, where predictive systems are implemented in ongoing operations. This would enable observation of how operators trust, behavior, and decision-making evolve over time, particularly in situations involving real risks. In this context, future studies could also explore how different types of explanations and visualizations influence operators trust and reliance on predictive systems. Future research could also investigate the role of prior experience with predictive systems, for instance by comparing operators with varying levels of familiarity, in order to better understand how experience influences trust and interaction. In addition, expanding the scope to include multiple industrial contexts and processes would support an assessment of the transferability of the guidelines developed in this study. Finally, combining qualitative findings with quantitative performance measures, such as response time, error rates, and process outcomes, may provide a more comprehensive evaluation of the impact of predictive

systems on operator performance.

8

Conclusion

This thesis set out to investigate how uncertainty in AI-based predictive systems can be effectively communicated to support decision-making in the process industry. Grounded in a human-centered design approach and conducted within the context of pulp and paper manufacturing, the work followed a Double Diamond process spanning qualitative work, iterative prototyping, and prototype evaluation with domain experts.

The findings show that operators mainly rely on experience-based and rule-based reasoning when dealing with uncertainty in their daily work. They interpret uncertainty through contextual cues, historical performance, and their understanding of process behavior. This highlights that uncertainty communication must align with existing mental models rather than introduce purely technical representations.

The results further demonstrate that the way uncertainty is communicated has a direct impact on trust, reliance, and decision-making. Transparent and contextualized uncertainty representations, such as explanations of predictions, confidence indicators, and access to historical accuracy, encourage more reflective decision-making and reduce the risk of over-reliance on AI systems.

Based on these insights, this thesis contributes a set of empirically grounded design guidelines for communicating uncertainty in AI-based predictive systems in the process industry. These guidelines emphasize the importance of contextual explanations, integration of historical performance, reduction of visual clutter, and alignment with operators workflows. The developed and evaluated prototype demonstrates how these principles can be implemented in practice, providing a concrete example of how uncertainty can be communicated in a way that supports operator understanding and decision-making.

This work contributes to the field of human-centered explainable AI by bridging the gap between theoretical principles and practical design in an industrial context. While previous research has often focused on general or experimental settings, this thesis provides insights grounded in real-world operator practices within a safety-critical domain.

However, the study is subject to limitations, including a limited sample size and evaluation within a specific industrial context. Future work could extend these findings by evaluating the proposed design guidelines across different industries, integrating them into real operational systems, and exploring long-term effects on

operator behavior and system trust.

In conclusion, effectively communicating uncertainty is not only a technical challenge but a design challenge. By aligning uncertainty communication with human cognition, work practices, and contextual needs, AI systems can better support informed decision-making and foster appropriate trust in industrial environments.

Bibliography

[1] F. Qian, Y. Tang, and X. Yu, "The future of process industry: A cyber-physical-social system perspective," *IEEE Transactions on Cybernetics*, vol. 54, no. 7, pp. 3878–3889, Jul. 2024. DOI: 10.1109/TCYB.2023.3298838.

[2] S. Prabhudesai, L. Yang, S. Asthana, X. Huan, Q. V. Liao, and N. Banovic, "Understanding uncertainty: How lay decision-makers perceive and interpret uncertainty in human-ai decision making," in *Proceedings of the 28th International Conference on Intelligent User Interfaces*, ser. IUI '23, Sydney, NSW, Australia: Association for Computing Machinery, 2023, pp. 379–396, ISBN: 9798400701061. DOI: 10.1145/3581641.3584033. [Online]. Available: https://doi.org/10.1145/3581641.3584033.

[3] M. Abdar, A. Khosravi, S. M. S. Islam, U. R. Acharya, and A. V. Vasilakos, "The need for quantification of uncertainty in artificial intelligence for clinical data analysis: Increasing the level of trust in the decision-making process," *IEEE Systems, Man, and Cybernetics Magazine*, vol. 8, no. 3, pp. 28–40, 2022. DOI: 10.1109/MSMC.2022.3150144.

[4] H.-J. Zimmermann, "An application-oriented view of modeling uncertainty," *European Journal of Operational Research*, vol. 122, no. 2, pp. 190–198, 2000, ISSN: 0377-2217. DOI: 10.1016/S0377-2217(99)00228-3. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0377221799002283.

[5] Y. Zhang, E. Brorsson, C. Westin, E. Zohrevandi, et al., "Human-centered explainable ai for process industries: Guidebook," EXPLAIN Project Consortium, Project Guidebook, 2025, ITEA4 Project (2022-2025). Contributors from ABB, Linköping University, Södra, Viking Analytics, and partners across Sweden, Germany, and the Netherlands.

[6] Swedish Forest Industries Federation, *Facts & figures*, Accessed: 2026-02-13, Dec. 2024. [Online]. Available: https://www.forestindustries.se/forest-industry/statistics/facts-and-figures/.

[7] Swedish Forest Industries Federation, *Snapshot of the swedish forest industry 2025*, Accessed: 2026-02-13, Nov. 2025. [Online]. Available: https://www.forestindustries.se/siteassets/bilder-och-dokument/rapporter/koll-pa-svensk-skogsindustri/snapshot-of-the-swedish-forest-industry-2025.pdf.

[8] Pappers, *Bransch*, Accessed: 2026-02-13. [Online]. Available: https://www.pappers.se/om-oss/bransch.

[9] T. Atkinson and M. Hollender, "Operator effectiveness," in *Collaborative Process Automation Systems*, M. Hollender, Ed., ISA (International Society of Automation), 2010, ch. 6.1, pp. 352–377, ISBN: 978-1-936007-10-3.

[10] H. M. K. Koskinen, P. Savioja, P. Mannonen, and M. Aikala, "The process is under control! understanding the building blocks of user experience in operator work," in *Proceedings of the 13th Nordic Conference on Human-Computer Interaction*, ser. NordiCHI '24, Uppsala, Sweden: Association for Computing Machinery, 2024, ISBN: 9798400709661. DOI: 10.1145/3679318.3685394. [Online]. Available: https://doi.org/10.1145/3679318.3685394.

[11] N. Flegel, J. Pöhler, K. Van Laerhoven, and T. Mentler, "i want my control room to be: On the aesthetics of interaction in a safety-critical working environment," Sep. 2022, pp. 488–492. DOI: 10.1145/3543758.3547562.

[12] J. Rasmussen, "Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-13, no. 3, pp. 257–266, 1983. DOI: 10.1109/TSMC.1983.6313160.

[13] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Psychological Review*, vol. 63, no. 2, pp. 81–97, 1956, ISSN: 1939-1471 (Electronic), 0033-295X (Print). DOI: 10.1037/h0043158.

[14] B. Liu, "In AI we trust? effects of agency locus and transparency on uncertainty reduction in human–AI interaction," *Journal of Computer-Mediated Communication*, vol. 26, no. 6, pp. 384–402, Nov. 2021. DOI: 10.1093/jcmc/zmab013. [Online]. Available: https://doi.org/10.1093/jcmc/zmab013.

[15] B. J. Zikmund-Fisher, D. M. Smith, P. A. Ubel, and A. Fagerlin, "Validation of the subjective numeracy scale: Effects of low numeracy on comprehension of risk communications and utility elicitations," *Medical Decision Making*, vol. 27, no. 5, pp. 663–671, 2007. DOI: 10.1177/0272989X07303824.

[16] U. Bhatt et al., "Uncertainty as a form of transparency: Measuring, communicating, and using uncertainty," in *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, ser. AIES '21, Virtual Event, USA: Association for Computing Machinery, 2021, pp. 401–413, ISBN: 9781450384735. DOI: 10.1145/3461702.3462571. [Online]. Available: https://doi.org/10.1145/3461702.3462571.

[17] E. Zohrevandi et al., "Design and visualisation of time-series based explanation mechanisms for industrial ai applications," in *Proceedings of the International Conference on Intelligence for Industry (IAI)*, Paper 73, 2025.

[18] J. Reyes, A. U. Batmaz, and M. Kersten-Oertel, "Trusting AI: Does uncertainty visualization affect decision-making?" *Frontiers in Computer Science*, vol. 7, 2025, ISSN: 2624-9898. DOI: 10.3389/fcomp.2025.1464348. [Online]. Available: https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2025.1464348.

[19] S. Cao, A. Liu, and C.-M. Huang, "Designing for appropriate reliance: The roles of ai uncertainty presentation, initial user decision, and user demographics in ai-assisted decision-making," *Proc. ACM Hum.-Comput. Interact.*, vol. 8,

no. CSCW1, Apr. 2024. DOI: 10.1145/3637318. [Online]. Available: https://doi.org/10.1145/3637318.

[20] A. Karagappa, P. K. Betz, J. Gilg, M. Zeumer, A. Gerndt, and B. Preim, "Enhancing uncertainty communication in time series predictions: Insights and recommendations," *ArXiv*, vol. abs/2408.12365, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:271924084.

[21] K. Ajani, E. Lee, C. Xiong, C. N. Knaflic, W. Kemper, and S. Franconeri, "Declutter and focus: Empirically evaluating design guidelines for effective data communication," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 10, pp. 3351–3364, 2022. DOI: 10.1109/TVCG.2021.3068337.

[22] M. Kay, T. Kola, J. R. Hullman, and S. A. Munson, "When (ish) is my bus? user-centered visualizations of uncertainty in everyday, mobile predictive systems," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16, San Jose, California, USA: Association for Computing Machinery, May 2016, pp. 5092–5103, ISBN: 9781450333627. DOI: 10.1145/2858036.2858558. [Online]. Available: https://doi.org/10.1145/2858036.2858558.

[23] L. Padilla, M. Kay, and J. Hullman, "Uncertainty visualization," in Feb. 2021, pp. 1–18, ISBN: 9781118445112. DOI: 10.1002/9781118445112.stat08296.

[24] A. Carter, S. Imtiaz, and G. Naterer, "Review of interpretable machine learning for process industries," *Process Safety and Environmental Protection*, vol. 170, pp. 647–659, 2023, ISSN: 0957-5820. DOI: https://doi.org/10.1016/j.psep.2022.12.018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957582022010837.

[25] K.-L. Du, R. Zhang, B. Jiang, J. Zeng, and J. Lu, "Understanding machine learning principles: Learning, inference, generalization, and computational learning theory," *Mathematics*, vol. 13, no. 3, 2025, ISSN: 2227-7390. DOI: 10.3390/math13030451. [Online]. Available: https://www.mdpi.com/2227-7390/13/3/451.

[26] R. Q. Majumder, "Machine learning for predictive analytics: Trends and future directions," *Available at SSRN*, Feb. 2025. DOI: 10.2139/ssrn.5267273. [Online]. Available: https://ssrn.com/abstract=5267273.

[27] J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cognitive Science*, vol. 12, no. 2, pp. 257–285, 1988. DOI: https://doi.org/10.1207/s15516709cog1202_4. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog1202_4. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1207/s15516709cog1202_4.

[28] C. Clark and R. Kimmons, "Cognitive load theory," *EdTechnica*, Jan. 2023. DOI: 10.59668/371.12980.

[29] N. A. Jones, H. Ross, T. Lynam, P. Perez, and A. Leitch, "Mental models: An interdisciplinary synthesis of theory and methods," *Ecology and Society*, vol. 16, no. 1, 2011, ISSN: 17083087. Accessed: Jan. 21, 2026. [Online]. Available: http://www.jstor.org/stable/26268859.

[30] C. Gonzalez, "Decision-making: A cognitive science perspective," in *The Oxford Handbook of Cognitive Science*. Oxford University Press, Oct. 2017, ISBN:

978-0-19-984219-3. DOI: 10.1093/oxfordhb/9780199842193.013.6. [Online]. Available: https://doi.org/10.1093/oxfordhb/9780199842193.013.6.

[31] C. D. Wickens, J. G. Hollands, S. Banbury, and R. Parasuraman, *Engineering Psychology and Human Performance*. London, UNITED STATES: Taylor & Francis Group, 2012, ISBN: 978-1-315-66517-7. [Online]. Available: http://ebookcentral.proquest.com/lib/chalmers/detail.action?docID=3570443.

[32] C. Vinney, *What is human-centered design? everything you need to know*, https://www.uxdesigninstitute.com/blog/what-is-human-centered-design/, Accessed: 2026-02-03, UX Design Institute, Nov. 2023.

[33] D. Doroftei et al., "User-centered design," in *Search and Rescue Robotics - From Theory to Practice*, London: IntechOpen, 2017, ch. 2. DOI: 10.5772/intechopen.69483. [Online]. Available: https://doi.org/10.5772/intechopen.69483.

[34] D. Norman and J. Nielsen. "The definition of user experience (ux)." Accessed: 2026-02-03, Nielsen Norman Group. [Online]. Available: https://www.nngroup.com/articles/definition-user-experience/.

[35] H. Sharp, J. Preece, and Y. Rogers, *Interaction Design: Beyond Human-Computer Interaction*. Newark, United States: John Wiley Sons, Incorporated, 2019, ISBN: 978-1-119-54735-8. [Online]. Available: http://ebookcentral.proquest.com/lib/chalmers/detail.action?docID=5746446.

[36] A. Pitale and A. Bhumgara, "Human computer interaction strategies designing the user interface," in *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2019, pp. 752–758. DOI: 10.1109/ICSSIT46314.2019.8987819.

[37] J. Simonsen and T. Robertson, Eds., *Routledge International Handbook of Participatory Design*, 1st ed. Routledge, 2012. DOI: 10.4324/9780203108543.

[38] T. Jiang, Z. Sun, S. Fu, and Y. Lv, "Human-ai interaction research agenda: A user-centered perspective," *Data and Information Management*, vol. 8, no. 4, p. 100 078, 2024, ISSN: 2543-9251. DOI: https://doi.org/10.1016/j.dim.2024.100078. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2543925124000147.

[39] R. Dwivedi et al., "Explainable ai (xai): Core ideas, techniques, and solutions," *ACM Comput. Surv.*, vol. 55, no. 9, Jan. 2023, ISSN: 0360-0300. DOI: 10.1145/3561048. [Online]. Available: https://doi.org/10.1145/3561048.

[40] G. Liu, "Uncertainty theory," *AI & SOCIETY*, Jul. 2025, ISSN: 1435-5655. DOI: 10.1007/s00146-025-02532-2. [Online]. Available: https://doi.org/10.1007/s00146-025-02532-2.

[41] Design Council, *The double diamond: A universally accepted depiction of the design process*, https://www.designcouncil.org.uk/our-resources/the-double-diamond/history-of-the-double-diamond/, Originally created in 2003 and launched publicly in 2004, 2004.

[42] J. Knopf, "Doing a literature review," *PS: Political Science Politics*, vol. 39, pp. 127–132, Feb. 2006. DOI: 10.1017/S1049096506060264.

[43] S. Hannabuss, "Research interviews," *New Library World*, vol. 97, no. 5, pp. 22–30, 1996, ISSN: included. DOI: 10.1108/03074809610122881. [Online]. Available: included.

[44] B. Hanington and B. Martin, *Universal Methods of Design, Expanded and Revised*. Beverly, UNITED STATES: Quarto Publishing Group USA, 2019, ISBN: 978-1-63159-749-7. [Online]. Available: http://ebookcentral.proquest.com/lib/chalmers/detail.action?docID=6039458.

[45] A. Bogner, B. Littig, and W. Menz, "Introduction: Expert interviews — an introduction to a new methodological debate," in *Interviewing Experts*, A. Bogner, B. Littig, and W. Menz, Eds. London: Palgrave Macmillan UK, 2009, pp. 1–13, ISBN: 978-0-230-24427-6. DOI: 10.1057/9780230244276_1. [Online]. Available: https://doi.org/10.1057/9780230244276_1.

[46] C. von Soest, "Why do we speak to experts? reviving the strength of the expert interview method," *Perspectives on Politics*, vol. 21, pp. 1–11, Jun. 2022. DOI: 10.1017/S1537592722001116.

[47] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006. DOI: 10.1191/1478088706qp063oa.

[48] K. Fu, M. Yang, and K. Wood, "Design principles: The foundation of design," Aug. 2015. DOI: 10.1115/DETC2015-46157.

[49] A. Cooper. "The origin of personas." Cooper Journal, Accessed: Feb. 4, 2026. [Online]. Available: https://cooper.com/journal/2008/05/the_origin_of_personas.

[50] R. I. Sutton and A. Hargadon, "Brainstorming groups in context: Effectiveness in a product design firm," *Administrative Science Quarterly*, vol. 41, no. 4, pp. 685–718, 1996, Cited by: 784. DOI: 10.2307/2393872. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0030354997&doi=10.2307%2f2393872&partnerID=40&md5=0ec818b35eb0ceb60c82eef2d0cd38ed.

[51] Google Design Sprint Kit, *Crazy 8s*, https://designsprintkit.withgoogle.com/methodology/phase3-sketch/crazy-8s, Accessed: 2026-03-11, n.d.

[52] Interaction Design Foundation. "What is how might we (hmw)?" Accessed: 2026-03-23, IxDF - Interaction Design Foundation. [Online]. Available: https://www.interaction-design.org/literature/topics/how-might-we.

[53] J. D. Van Der Laan, A. Heino, and D. De Waard, "A simple procedure for the assessment of acceptance of advanced transport telematics," *Transportation Research Part C: Emerging Technologies*, vol. 5, no. 1, pp. 1–10, 1997.

[54] M. Crouch and H. Mckenzie, "The logic of small samples in interview-based," *Social Science Information Sur Les Sciences Sociales - SOC SCI INFORM*, vol. 45, pp. 483–499, Dec. 2006. DOI: 10.1177/0539018406069584.

[55] Figma, Inc., *Figma design software*, 2026. [Online]. Available: https://www.figma.com.

[56] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, DOI: 10.1109/MCSE.2007.55.

[57] Anthropic, *Claude*, https://claude.ai/, Large language model, 2026.

[58] Lovable Labs, *Lovable*, https://lovable.dev/, AI development platform, 2026.

[59] X. Dai, M. T. Keane, L. Shalloo, E. Ruelle, and R. M. Byrne, "Counterfactual explanations for prediction and diagnosis in xai," in *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, ser. AIES '22, Oxford, United Kingdom: Association for Computing Machinery, 2022, pp. 215–226, ISBN: 9781450392471. DOI: 10.1145/3514094.3534144. [Online]. Available: https://doi.org/10.1145/3514094.3534144.

[60] J. Wexler, M. Pushkarna, T. Bolukbasi, M. Wattenberg, F. Viégas, and J. Wilson, "The what-if tool: Interactive probing of machine learning models," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 1, pp. 56–65, 2020. DOI: 10.1109/TVCG.2019.2934619.

[61] D. Leslie, "Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of ai systems in the public sector," en, *The Alan Turing Institute*, 2019. DOI: 10.5281/ZENODO.3240529. [Online]. Available: https://zenodo.org/record/3240529.

[62] European Parliament. "Eu ai act: First regulation on artificial intelligence, The first comprehensive ai law." Accessed: 2026-02-03, European Parliament, Accessed: Feb. 3, 2026. [Online]. Available: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

[63] OpenAI, *Chatgpt*, https://chat.openai.com/chat, Large language model, 2026.

A

Appendix 1

A.1 Initial Interview Questions

Demographics

1. Can you tell me about your role/title here at Braviken?
2. How long have you worked as an operator?
3. What education/background do you have?
4. What is your age?
5. What gender do you identify as?

Work Context

1. Describe a typical shift. What do you do?
2. Which systems/tools do you use daily?

Uncertainty

1. Can you tell us about what kind of decisions you make during your work?
   (a) How much of your decisions is guided by clear rules and procedures, and how much relies on your own judgement?
   (b) How do you experience the precision and usefulness of these rules in your work?
   (c) Do you sometimes deviate from the rules and rely on your own judgement?
2. Can you give an example of a moment when something wasnt fully clear, but you still had to act or decide?
   (a) What was not fully clear in that situation?
   (b) How did you reason your way through it?
   (c) Do you sometimes have to make assumptions?

(d) How common is it to be in an unclear situation?

(e) How did you feel in that situation?

Distributed Control System (DCS)

1. How do you usually get an overview of the process in the control room?
   (a) How are the different processes visualized in the DCS?
   (b) How easy is it for you to understand the visualizations in DCS?
   (c) Are there any visualizations you find difficult to interpret? Why?
2. Is there any information you wish you had on the DCS screens to make your job easier?

The AI Concept

1. Suppose there is a predictive system that provides predictions about when brightness will increase. What would you need to know about the prediction to feel confident using it?
   (a) Would you want to know the accuracy of the prediction?
   (b) Would you want to know what the system bases its prediction on?
   (c) Would you want to know how often the predictive system has been correct in the past?
   (d) Would you trust such a system?
2. What would this system have to do for you to consider it as a good teammate, versus a distraction?
   (a) Why? Why not?
   (b) What would make you consider it as a good teammate?
   (c) If the predictions were uncertain, would your trust be affected?
   (d) Would you want the system to tell you how certain or uncertain its predictions are?

A.2 Evaluation

Semi-structured Interview Questions

1. Describe what you see on the screen.
2. Do you find any function or visualization particularly helpful?
3. When you look at the uncertainty visualization (e.g., the graph or percentage indicator) - how do you interpret it?

4. What does '42% moderate confidence' mean to you in practice?
5. How does the way uncertainty is presented affect your trust in the system's predictions?
6. Would you be willing to act on this prediction in real life? Why or why not?
7. How would you weigh the system's prediction against your own experience?
8. How would this information influence your decisions about the bleaching process?
9. Is there anything you feel is missing to make a confident decision?

A.3 Van der Laan Acceptance Scale

How do you perceive this visualization of AI uncertainty?

1.	Useful	|_|_|_|_|_|	Useless
2.	Pleasant	|_|_|_|_|_|	Unpleasant
3.	Bad	|_|_|_|_|_|	Good
4.	Nice	|_|_|_|_|_|	Annoying
5.	Effective	|_|_|_|_|_|	Superfluous
6.	Irritating	|_|_|_|_|_|	Likeable
7.	Assisting	|_|_|_|_|_|	Worthless
8.	Undesirable	|_|_|_|_|_|	Desirable
9.	Raising Alertness	|_|_|_|_|_|	Sleep-inducing
10.	Reliable	|_|_|_|_|_|	Unreliable
11.	Easy to understand	|_|_|_|_|_|	Hard to understand
12.	Clear	|_|_|_|_|_|	Unclear

A.4 Consent Form

Informed consent

Thank you for agreeing to participate in our research project on **'Designing for uncertainty in AI'** and human automation interaction. The project investigates how operators in the pulp and paper industry interpret and manage uncertainty in their work, particularly in relation to AI prediction systems. The goal is to develop design guidelines and prototypes for communicating uncertainty in AI outputs in ways that support operators' trust and decision-making. At ABB, we are exploring how these advancements unlock new efficiencies, empower workers, and reshape decision-making. By identifying key motivators, assessing technical capabilities, and harnessing domain expertise, we aim to create future user experience (UX) that enhance both productivity and human potential.

During our interview & evaluation, we will gather insights and inputs from you through conversations, which will be used internally and ethically. Your privacy and comfort are our highest priorities. Participation is voluntary, and you are free to refuse to answer questions or stop the discussion at any point if you feel uncomfortable.

In compliance with GDPR, we request your consent to retain certain personal information. Below is a list of the personal information that may be retained, its purpose, and the corresponding duration for which this data will be saved. All content generated by you (digital, oral, or written) will be **anonymized** before being used in further analysis and/or development.

S.No	Personal Data	How will the data be used	Duration
1	First name and Last name	- To initially record our observations - As part of our raw data records. This information is held confidentially within the research team.	1 year
2	Organization	To keep track of our partners in this research project	8 years
3	Interview notes	The researchers will take notes during the interviews to ensure all information is gathered.	6 months
4	Photographs	Photos might be taken during the study for documenting purposes. These pictures may be used internally as a part of explaining the work and its results.	8 years

☐ **I have read the above and agree to the terms of usage of my personal information.**

Name_____ Signature: _____

For further questions or changes in preferences, please contact
Lisa Ljungberg, Chalmers University of Technology (lisalju@chalmers.se)
Supervisor: Mowei Yang (mowei.yang@se.abb.com)at ABB Corporate Research, Sweden