Guidebook

EXPLAIN Project 2022-2025

Human-Centered Explainable AI for Process Industries

How to design and develop interactive Explainable AI (XAI) for process industries

Table of contents

01 – 02	Chapter 1	**Introduction**
03 – 05	Chapter 2	**From Explainable AI (XAI) to Human-Centered Explainable AI (HCXAI)**
06	Chapter 3	**Market Landscape**
07	Chapter 4	**Process Industries**
08 – 11	Chapter 5	**User-Centered Design Approach for Developing Industrial XAI Solutions**
12 – 26	Chapter 6	**How to Design and Develop Human-Centered Explanation and Feedback for Industrial AI Solutions**
27 – 28	Chapter 7	**Human-Centered MLOps for Trustworthy Industrial AI**
29 – 30	Chapter 8	**Turning Concepts to Real Deployment in Industries: Business Perspectives**
31	Chapter 9	**Conclusion and Future Outlook**
32	Chapter 10	**References**

1. Introduction

This guidebook presents key insights and lessons-learned from the EU-funded project "EXPLAIN (EXPLanatory interactive Artificial intelligence for INdustry)." It investigates how to design and develop Human-Centered Explainable AI for process industries, using real-world examples to showcase its impact.

Artificial Intelligence (AI) has rapidly reshaped industries across the globe, revolutionizing operations, improving efficiency, and minimizing waste. Among the sectors benefiting from this transformation is the process industry—a diverse field that includes oil and gas, chemicals, pulp and paper, metals, cement, food, and beverages. These industries rely on intricate production processes such as blending, chemical reactions, and refining, all of which must meet rigorous standards for quality, safety, and efficiency. While automation plays a significant role in these industries, human expertise remains indispensable. Operators, engineers, and maintenance teams ensure systems run smoothly, adapt to changes, and respond to unforeseen challenges. Together, they form the backbone of operations, even in an era of increasing reliance on Artificial Intelligence.

Yet, these powerful systems (AI) often function as "black boxes", executing highly complex calculations without revealing how they reach their conclusions. This lack of transparency presents serious obstacles:

- **Trust**: Operators may hesitate to rely on AI systems they cannot understand;
- **Usability**: Without clear explanations, AI recommendations can be misinterpreted, misused, or ignored;
- **Adoption**: A lack of trust and transparency hinders the broader integration of AI into day-to-day operations.

EU AI Act

- **Article 13** states that "*High-risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately*" [1].
- **Article 14** states that systems must be designed and developed to allow "*effective oversight by natural persons during the period in which the AI system is in use*." It further specifies that users must be enabled "*to correctly interpret the high-risk AI system's output, taking into account, for example, the interpretation tools and methods available*." [1].

[1] (EU AI Act: First Regulation on Artificial Intelligence | Topics | European Parliament, 2023)

In addition, the upcoming EU AI Act introduces specific requirements for explainability in high-risk AI applications—many of which are directly relevant to industrial use cases involving safety. It requires the system to be transparent and also ensure human oversight (see Article 13-14)

Explainable AI (XAI) therefore plays a critical role in this legal framework. XAI aims to make AI systems interpretable and understandable, building a bridge between advanced machine intelligence and the human trust required for effective collaboration. In doing so, it transforms the relationship between operators and technology, enabling them to work with automation rather than around it.

XAI makes AI systems more understandable, helping build the trust needed for effective human-AI collaboration. However, while progress has been made in XAI for many fields, process industries pose unique challenges. These systems not only handle massive amounts of data but must also account for time-series patterns, where the sequence and timing of events are as critical as the data itself. For AI to truly support decision-making in these environments and get adopted by users in industries, it must provide not just accurate output, but also clear, actionable, and relevant explanations tailored to the domain.

Despite growing demand across industries for greater explainability, there is still a lack of practical guidance on how to develop XAI solutions in real-world industrial settings. Practitioners need actionable strategies to address the constraints and realities of operational environments.

This guidebook seeks to bridge that gap. It offers practical insights, best practices, and lessons learned from real-world design and deployment of XAI solutions. It is intended to support AI developers, engineers, and decision-makers in the process industry as they work to integrate explainability into their systems. By sharing concrete examples and industry-driven insights, our goal is to help organizations make AI-driven decisions that are more transparent, trustworthy, and effective—ultimately enhancing efficiency, safety, and human trust in AI-powered operations.

1.2 Key Findings

The guidebook draws on key findings from the EU-funded project EXPLAIN, which focuses on developing Human-Centered Explainable AI solutions tailored to the process industries. Here are the highlights:

01

Social, Interactive, and Explorative Explanations

Explanations should not be purely technical but also social, interactive, and explorative to enhance user understanding and engagement. The right explainability approach depends on the specific needs of the user (explainee) and the context in which AI is applied.

02

Alignment with Mental Models of Users

- Understanding causal relationships is crucial in many industrial use cases.
- XAI solutions should be designed to align with the mental models of users to improve trust, usability, and decision-making.

03

Explainability and Trust

Explainability can serve as a trust calibration tool. However, trust extends beyond technical explanations—it is also shaped by the ability of enabling human control over AI as well as long term usefulness of AI solutions for the workers in their daily tasks.

04

Challenges in Feedback Implementation

- Implementing feedback mechanisms is difficult due to retraining cycles that can — for complex modes — take hours or days.
- Feedback effectiveness is directly linked to the quality and clarity of explanations.

05

User-Centered Design Approach is Key

- XAI development should follow a user-centered process, incorporating user needs, design, evaluation, and iteration, supported by strong collaboration.
- Evaluating explanations through a mixture of methods is effective.

06

Technical Readiness in Industries is Crucial for any AI Development

Developing AI solutions requires a foundational level of technical readiness. XAI introduces additional complexity, making scalable solutions and MLOps integration essential.

07

Develop Scalable and Infrastructure-Aware Solutions

Due to high computational demands and the unique constraints of each use case, implementing XAI in industrial settings requires scalable solutions and adaptable MLOps infrastructure that balance on-premises and cloud performance.

08

Collaboration Across Stakeholders

Effective XAI implementation requires close collaboration among key stakeholders, including AI developers, domain experts, and operators, to ensure practical and meaningful solutions.

09

Distinguish AI Performance from Explanation Quality

AI model performance and explanation accuracy should be evaluated separately. While designers see the AI and XAI models as distinct, operators often don't—posing a design challenge.

All these insights will be explained in more detail in the following chapters. EXPLAIN project is introduced on Page 4.

Background and EXPLAIN Project

2. From Explainable AI (XAI) to Human-Centered Explainable AI (HCXAI)

Artificial Intelligence is transforming many industries at an unprecedented pace, and process industries are no exception. However, as AI systems grow more complex, understanding how they make decisions becomes both crucial and challenging. This need gave rise to Explainable AI (XAI)—a set of techniques and methodologies aimed at making AI decisions transparent and understandable.

An explanation offers an "*interface between humans and a decision maker that is both an accurate proxy of the decision maker and comprehensible to humans*" [2]. Current explainability methods are broadly categorized into two types:

- Ante-hoc Explanations: Built-in transparency within simple models, such as decision trees or linear regressions, which are inherently interpretable.
- Post-hoc Explanations: External tools applied to interpret complex models after they have been developed, such as feature importance. In the EXPLAIN project, we have focused on this type of methods.

However, explainability is not solely a property of the model itself—it depends on how the person receiving the explanation perceives and understands it [2]. Simply making a model fully transparent does not ensure that users can interpret or apply the information effectively. The abovementioned known explainability approaches help improve technical transparency, but they often fall short in addressing the broader spectrum of stakeholder needs.

[2] Guidotti, R., Monreale, A., Ruggieri, S., Pedreschi, D., Turini, F., & Giannotti, F. (2018). Local rule-based explanations of black box decision systems. arXiv preprint arXiv:1805.10820. https://doi.org/10.48550/arXiv.1805.10820

Most explanations are optimized for technical experts, leaving domain specialists, operators, and decision-makers in process industries with limited insights.

A good explanation should provide relevant, comprehensible, and actionable insights, which depend on factors such as the user's prior knowledge, goals, and cognitive load. Therefore, creating effective XAI solutions requires a human-centered approach that prioritizes users' explainability needs and defines success based on their experience, understanding, and empowerment.

Recognizing that AI operates in human environments and the needs of more stakeholders should be considered, researchers have begun to emphasize Human-Centered XAI (HCXAI). As a result, a growing research community in human-centered XAI is incorporating cognitive, sociotechnical, and design perspectives to improve AI's usability and impact. This paradigm shifts focus from merely providing explanations to designing AI systems that inherently prioritize human needs, trust, and usability.

Process industries, such as chemical, energy, and manufacturing sectors, operate in environments where decision-making is high-stakes and time-sensitive. The integration of HCXAI principles in such industries enhances operational efficiency, safety, and compliance.

"There is no one-fits-all solution in the growing collection of XAI techniques. The technical choices should be driven by users' explainability needs."

[3] Liao, Q. V., & Varshney, K. R. (2021). Human-centered explainable ai (xai): From algorithms to user experiences. *arXiv preprint* arXiv:2110.10790. https://doi.org/10.48550/arXiv.2110.10790

Transitioning to HCXAI is not just about meeting regulatory standards—it should be a strategic move to unlock AI's full potential. Organizations that prioritize human-centered design in XAI systems report greater adoption, improved performance, and stronger stakeholder trust.

According to McKinsey trend reports, there are clear benefits of Explainable AI (XAI) for positive Return of Investmant (ROI) [4]:

- **Operational-Risk Mitigation** - XAI reduces risks by identifying potential issues like bias or inaccuracy early, enabling organizations to fine-tune AI systems and introduce oversight, particularly in high-stakes areas like fraud detection.
- **Regulatory Compliance and Safety** - Ensures AI systems operate within legal, ethical, and regulatory frameworks, protecting against penalties and biases, especially in sensitive applications like hiring.
- **Continuous Improvement** - Supports debugging and iterative refinement by offering insights into AI system behavior, helping maintain alignment with business goals, such as improving recommendation engines.
- **Stakeholder Confidence** - Builds trust by enhancing transparency, enabling users—like healthcare professionals—to understand AI outputs, boosting adoption and reliability.
- **User Adoption** - Increases user satisfaction and adoption by aligning AI outputs with expectations, driving innovation and growth.

[4] Giovine, C., & Roberts, R. (2024, November 26). Building AI trust: The key role of explainability. McKinsey & Company. Building AI trust: The key role of explainability | McKinsey

In sum, by embedding human-centered principles into XAI design and governance, industries can not only enhance operational efficiency but also foster a collaborative relationship between humans and machines.

EXPLAIN-EXPLanatory interactive Artificial intelligence for INdustry

EXPLAIN, short for EXPLanatory interactive Artificial intelligence for INdustry, envisions a future where AI is not just intelligent but also transparent, interactive and understandable. AI holds vast potential for improving industrial operations, from optimizing processes to reducing waste. Yet, its success hinges on trust. EXPLAIN aims to bridge this gap by ensuring transparency throughout the AI development cycle. Stakeholders can track how decisions are made, uncover biases, and make informed adjustments. This transparency fosters trust and paves the way for AI systems to handle tasks where precision and accountability are paramount.

- EU funded project (ITEA)
- 15 Partners & 8 Use Cases
- Runtime: May 2022 – April 2025
- Financial Support:

VINNOVA Sweden (2021-04336)
Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR; 01IS22030)
Rijksdienst voor Ondernemend Nederland (AI212001)

In this project, we propose an end-to-end Machine Learning lifecycle by integrating stakeholders—ML experts, domain specialists, and end-users—throughout the process. This collaborative method unfolds in three key phases:

- **Explanatory Training:** During this phase, ML experts and domain specialists interact directly with AI models. Explanations accompany outputs, shedding light on how decisions are made. Feedback loops allow for iterative improvements, enabling the model to learn and correct itself in real time.

- **Explanation Review:** Once training is complete, domain experts evaluate the model's internal reasoning. Does it capture the right concepts from the data? Has it learned any misleading biases? These checks help refine the model further and ensure it aligns with real-world expectations.

- **Interactive Validation and Feedback:** After deployment, end-users engage with the AI through visual dashboards and interactive tools. They can interpret results, provide feedback, and even trigger incremental training to refine predictions and outputs over time.

By incorporating these steps into a unified MLOps (Machine Learning Operations) framework, EXPLAIN prevents the disconnection between ML developers and domain experts — a common issue that often slows AI adoption.

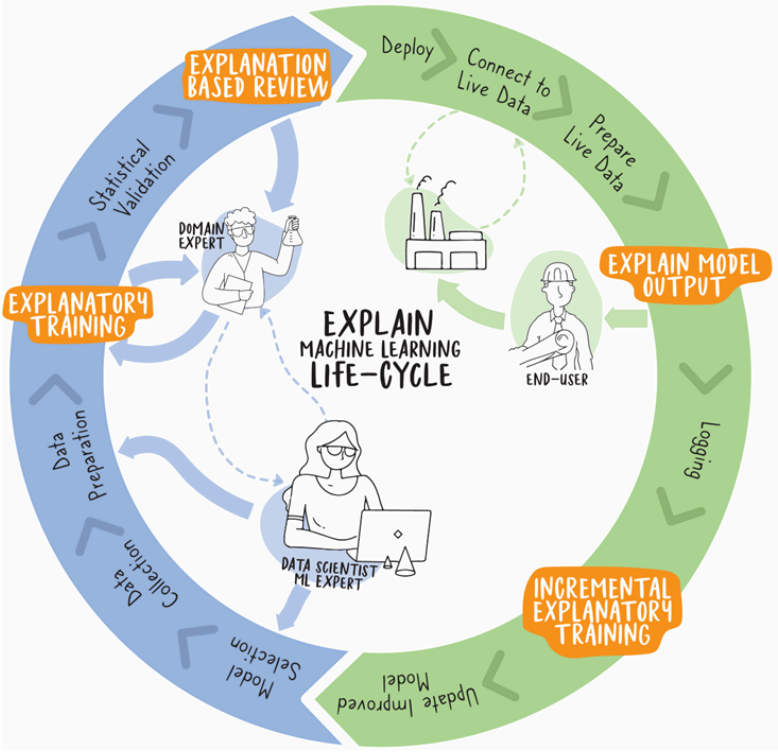


Figure 1 The end to end ML lifecycle proposed in EXPLAIN

A core feature of EXPLAIN's approach is its use of visualizations and interactive dashboards. These tools make AI insights accessible and actionable, empowering users to explore model outputs, test scenarios, and identify weaknesses. Feedback mechanisms allow these insights to refine predictions and enhance future performance, creating a continuous improvement loop.

As industries worldwide face the challenge of AI adoption, EXPLAIN offers a human-centered solution that promises to transform skepticism into confidence and potential into progress.

With partnerships across Germany, the Netherlands, and Sweden, EXPLAIN exemplifies the power of collaboration in tackling the most pressing challenges in AI. It's a step toward a future where AI doesn't just work for us but works with us—empowering people, enhancing industries, and building a smarter, more sustainable world.

3. Market Landscape

Heavy industries are experimenting with AI, and major companies are making significant investments to integrate these technologies into their products. In turn comes a growing demand for transparency, trust, and regulatory compliance, which further accelerates the development of XAI solutions.

Several major players are shaping the XAI landscape:

- **Amazon:** SageMaker Clarify provides tools for ML modelers and developers to understand model characteristics and debug predictions, positioning XAI as a technology differentiator.
- **Google:** Vertex AI includes feature-based explanations and example-based explanations to improve model interpretability.
- **IBM:** Watsonx.governance enhances transparency across the AI lifecycle with integrated explainability techniques.
- **Microsoft:** Azure Responsible AI dashboard with explainability techniques such as SHAP, LIME, counterfactual analysis and causal inference.

In 2023, the global XAI market was valued at approximately $6.2 billion and is projected to reach between $16.2 billion and $39.6 billion by 2028, reflecting a compound annual growth rate (CAGR) of around 20% during the forecast period [5]. Companies such as Google, IBM, Microsoft, and Amazon, offer platforms with integrated general XAI techniques. However, these platforms are often developed without industry-specific expertise, making it difficult to tailor XAI solutions to the users in industry and their particular needs.

[5] Market and Markets. (2025, May 7). *Explainable AI Market worth $16.2 billion by 2028.* https://www.marketsandmarkets.com/PressReleases/explainable-ai.asp

[6] Di Bonito, L. P., Campanile, L., Di Natale, F., Mastroianni, M., & Iacono, M. (2024). eXplainable Artificial Intelligence in Process Engineering: Promises, Facts, and Current Limitations. *Applied System Innovation* (ASI), 7(6). doi: 10.3390/asi7060121.

When examining companies that offer AI applications specifically for industry, a different trend emerges regarding XAI. These solutions, such as those created by Trendminer and Seeq, tend to prioritize usability and accessibility but often lack features or techniques focused on explainability.

- **Trendminer MLHub:** Offers a no-code platform for analysis, monitoring, anomaly detection, and prediction, using machine learning across various use cases. While offering accessibility through a user-friendly no-code interface, it does not highlight any specific XAI techniques.
- **Seeq:** Helps users analyze process data but primarily enhances usability rather than model transparency.

Although usability and accessibility are important components of explainability they do not fully address the fundamental concerns of using AI in industrial contexts, such as trust in AI-driven decision-making and regulatory compliance.

Scientific research into applied XAI solutions has been limited by a lack of large, reliable datasets [6], and while big tech firms focus on general XAI frameworks and explainability techniques, smaller companies develop AI solutions tailored to specific industries, often without robust explainability features.

This gap presents an opportunity for industry leaders—who have greater resources, access to data, and expertise in industrial domains—to take the lead in developing practical XAI applications tailored to the needs of industry users. The work and insights presented in this guidebook are concrete steps in that direction.

4. Process Industries

Process industries, including sectors such as chemical manufacturing, oil and gas refining, food processing, pharmaceuticals, and pulp and paper, are characterized by complex, continuous operations that require specialized systems, stringent safety measures, and robust process control.

Unlike discrete manufacturing, where individual items are produced, process industries often operate in a continuous or semi-continuous manner. Their processes are integrated systems where every component—from raw material handling to the final packaging—must be carefully managed.

Key characteristics of process industries:

- **Continuous Operations:** Many process plants run 24/7, requiring round-the-clock monitoring and control.
- **Complex Integration:** The production process typically involves numerous interconnected subsystems such as reactors, separators, heat exchangers, and storage facilities.
- **Safety and Environmental Compliance:** With high-energy operations and hazardous chemicals, safety protocols and environmental regulations are fundamental.
- **Capital-Intensive Investments:** Establishing a process plant demands significant capital outlay, sophisticated technology, and long-term planning.

In this context, process industries are designed not only for efficiency and productivity but also for reliability and resilience, ensuring that disruptions are minimized, and operational risks are well-managed.

The uniqueness of process industries lies in their operational complexity, the need for continuous production, and the integration of advanced control systems.

Images generated by Google Gemini

Time series data
Time series data is at the heart of process industry operations. Sensors and control systems generate massive amounts of data that reflect the performance and health of every process variable.

Key characteristics of time series data:

- **High frequency and volume:** Data is collected every few seconds or milliseconds, creating large datasets.
- **Temporal dependency:** Values depend on previous points, making trends and patterns essential for forecasting.
- **Multivariate nature:** Many variables (e.g., temperature, pressure, flow) are tracked together, requiring holistic analysis.
- **Anomaly detection:** Continuous monitoring helps catch malfunctions or early signs of failure.
- **Historical insight:** Past data supports optimization, troubleshooting, and predictive maintenance.

Effective time series management ensures safe, efficient, and reliable operations.

Sorting through vast amounts of data to detect issues or predict failures is a major challenge for human operators. With growing complexity, human operators face challenges in identifying issues or predicting failures—this is where Machine Learning (ML) steps in. ML models can detect patterns and forecast problems, but their "black box" nature can limit trust. Explainable AI (XAI) tackles this by making model decisions more transparent and easier for humans to understand.

5. User-Centered Design Approach for Developing Industrial XAI Solutions

To ensure XAI solutions align with user needs and real-world challenges, a *User-Centered Design (UCD)* approach is essential. User-centered design is an approach that puts end-user's needs, expectations, and problems at the forefront of the design process.

User-Centered Design Cycle

The process can be broken down into three key phases:

Phase 1: Understanding Requirements for Explainability (& Feedback)

- Conduct field studies to explore user workflows, challenges, and decision-making needs.
- Identify what Machine Learning should do to tackle the user challenges
- Analyze insights from user interactions with to define explainability user requirements and for explainer and feedback mechanisms.
- Capture domain-specific requirements to ensure AI explanations are contextually relevant and actionable.
- Identify challenges which end users face in interpreting the AI model's output.

Phase 2: Developing Prototypes that Contain Explainers (& Feedback)

- Build use-case-specific prototypes incorporating AI explainability features.
- Ensure both front-end (UI/UX) and back-end (model explanations) are designed to align with user expectations.
- Iterate on designs to balance usability, accuracy, and transparency in AI-generated explanations.

Phase 3: Evaluating Prototypes with End Users

- Conduct direct user testing in real-world settings, such as customer sites.
- Gather feedback to refine AI-generated explanations, improving clarity and trust.
- Adapt explainability features based on user cognitive load, domain expertise, and task complexity.

This approach will make sure users needs are systematically gathered, translated to design and iteratively improved through validation. By following this iterative, user-focused process, XAI solutions could be made more intuitive, trustworthy, and effective in supporting human AI collaborations in process industries.

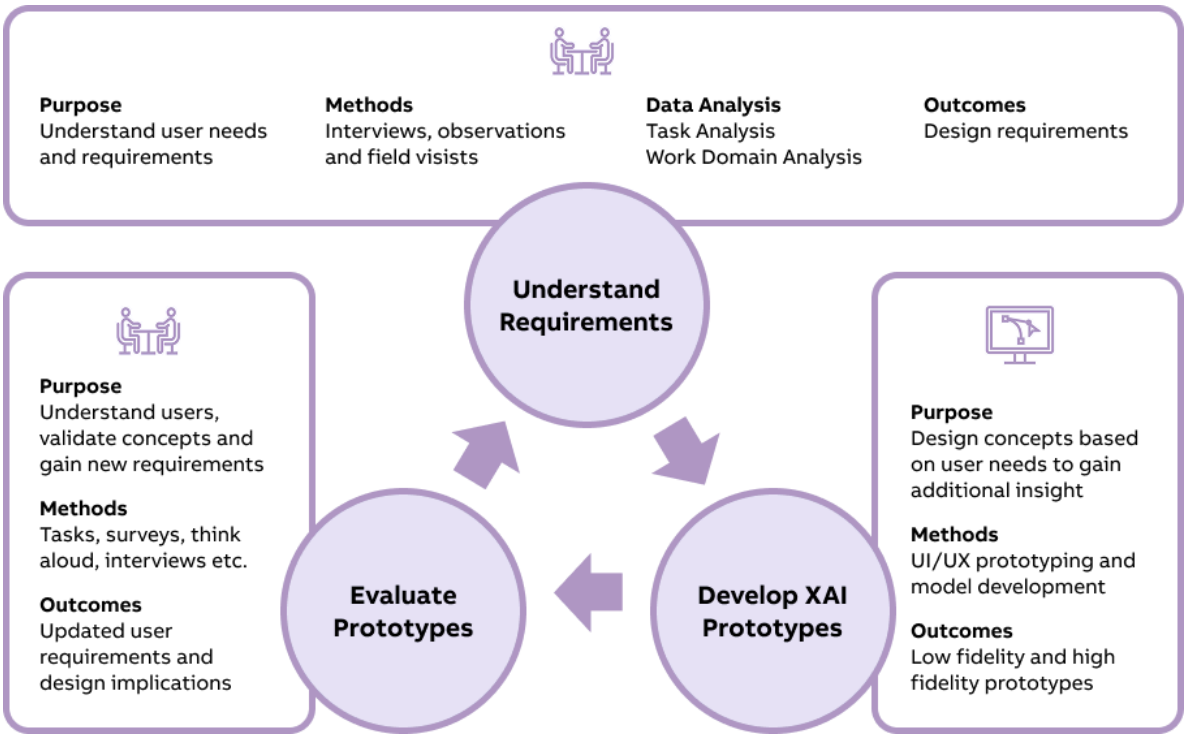


Figure 2 User-Centered Design Cycle

Gathering User Requirements for XAI solutions

Since mature AI solutions have not yet been widely deployed in industrial settings, gathering user requirements remains a challenge. To develop AI solutions that are truly useful, it is essential to understand users' tasks and goals, current workflows, interactions with existing automated systems, the challenges they face, and their attitudes and expectations toward AI. Based on this understanding, we can identify the roles machine learning should play in supporting their work, as well as the level and type of explainability needed for those solutions.

Methods for effectively gathering and analyzing XAI requirements and needs

Task Analysis

Task analysis connects user interactions, information needs, and goals to the specific stages of the industrial process that operators are managing [7]. By mapping out these relationships, it becomes possible to derive context-specific requirements for explainability.

[7] Hackos, J. T., & Redish, J. C. (1998). *User and task analysis for interface design*. John Wiley & Sons, Inc..

In industrial settings, the relevance and nature of AI-generated suggestions often vary depending on the stage of the process they pertain to. This analysis ensures that explainability is not treated as a one-size-fits-all feature, but rather as something tailored to the distinct demands of each process stage. It supports the development of targeted explanations that align with operator goals and information requirements, ultimately leading to more effective and trustworthy human-AI collaboration.

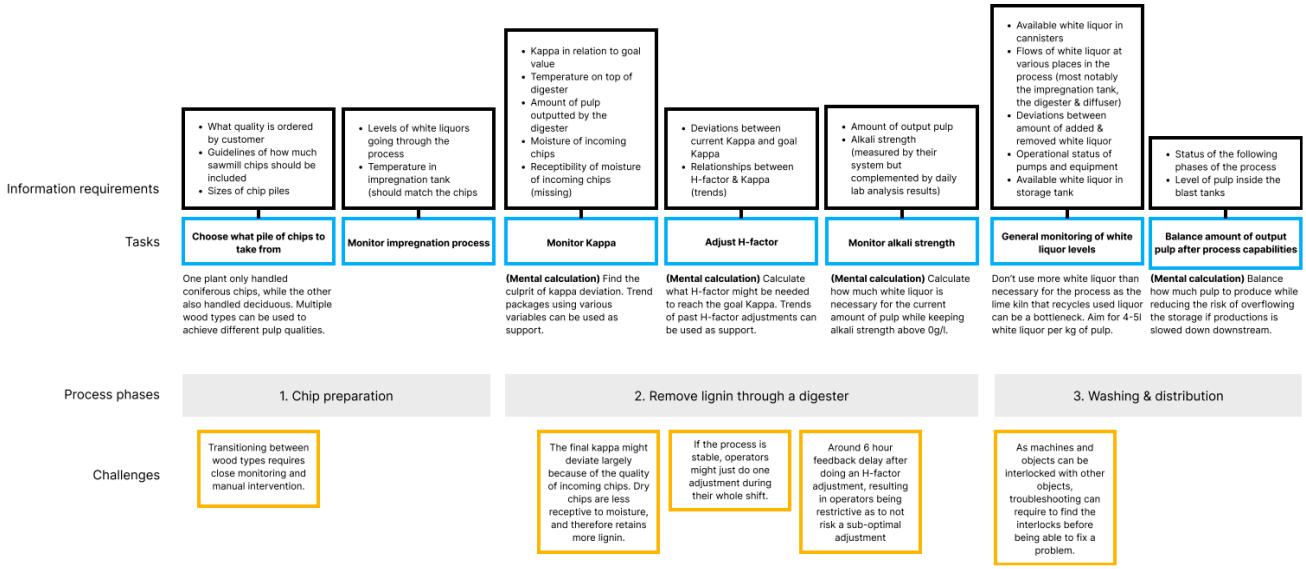


Figure 3 High level task analysis for the digester use case in a Pulp mill

Work Domain Analysis

Work Domain Analysis (WDA) is a task abstraction technique designed to model the complexity of a work environment by defining the boundaries of the solution space in which operators' function [8]. It does this by identifying hierarchical means-end relationships among tasks, enabling a deep understanding of the domain's functional structure. Unlike conventional task abstraction approaches that define tasks based on user needs, WDA derives tasks from system functions that must be supported, ensuring operator behavior remains within safety and efficiency boundaries. This makes WDA particularly robust for designing interfaces in complex domains, as it avoids inefficiencies caused by ill-defined user tasks.

In this project, WDA was used to analyze several use cases [9], guiding the design of an interface that enhances operator performance by aligning system functionality with domain-specific tasks.

Frame Requirements as Questions

The identified requirements were translated into questions that users might ask about the models they interact with. This approach, inspired by Liao et al. [10], helps to better capture user needs in terms of explainability. Framing explainability requirements as questions provides greater granularity, highlighting which aspects are most important to users. This makes it a valuable method for guiding model development and aligning with the specific information needs of users in a given context.

[8] Naikar, N. (2016). *Work domain analysis: Concepts, guidelines, and cases*. CRC press.

[9] Zohrevandi, E., Brorsson, E., Darnell, A., Bång, M., Lundberg, J., Ynnerman, A. (2023). Design of an Ecological Visual Analytics Interface for Operators of Time-Constant Processes. *2023 IEEE Visualization and Visual Analytics (VIS)*, Melbourne, Australia, 2023, pp. 131-135, doi: 10.1109/VIS54172.2023.00035.

[10] Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: informing design practices for explainable AI user experiences. *In Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-15). https://doi.org/10.1145/3313831.3376590

Evaluation of XAI concepts and prototypes

The evaluation of XAI concepts or prototypes is a complex and challenging task. Existing methods often focus on technical properties of explanations without considering human factors, but the needs and expectations for how a model should be explained does differ on the context and receiver of the explanation, therefore making human factors essential for developing human-centered XAI solutions. By conducting user evaluations, we can gain valuable insights that significantly enhance acceptance, trust, decision support, user experience, and overall usability. This section summarizes key insights and recommendations for designing and conducting effective user evaluations tailored to process industries.

Importance of User Evaluation
User evaluation is a crucial step in developing XAI solutions that truly support human decision-making. Explainability can be highly contextual and individual depending on users' role, domain expertise and technical fluency. Throughout various evaluations, we have observed that even small UX improvements can make a substantial difference. Engaging users early and frequently, even if only briefly, ensures that the solutions align with their needs, making AI explanations more comprehensible and actionable.

Designing Suitable Methods for Different Development Phases
The level of detail (fidelity) in your prototype should guide your choice of evaluation method. Early in development, when exploring many potential directions, use simple, low-fidelity prototypes. This allows for faster, cheaper evaluation of fundamental usability aspects without investing heavily in designs that might change. As the project progresses and designs become more defined, shift to higher-fidelity prototypes. These support more structured, realistic experiments to test specific user interactions and comprehension effectively. Ensure methods match fidelity; complex task evaluations require realistic high-fidelity prototypes to generate meaningful insights.

Defining the Purpose of Evaluation
Before conducting user evaluations, it is essential to establish clear objectives, as different evaluations serve different purposes. If the goal is to assess how well users understand model outputs, evaluation methods should be tailored to capture and measure user understanding. If the goal is to explore different approaches of XAI on user acceptance, evaluations methods should be created to measure differences in acceptance between different prototypes, and so on. For more examples of evaluation goals and suitable methods to match the goal, we recommend work done by Nauta et al. [11]. Keep in mind that operators in process industries tend to focus on process-related discussions, which may divert attention from evaluating ML output understanding. Facilitators should guide them back to the evaluation goals.

[11] Nauta, M., Trienes, J., Pathak, S., Nguyen, E., Peters, M., Schmitt, Y., ... & Seifert, C. (2023). From anecdotal evidence to quantitative evaluation methods: A systematic review on evaluating explainable ai. *ACM Computing Surveys, 55*(13s), 1-42. https://doi.org/10.1145/3583558

Combining Evaluation Methods for XAI
To effectively evaluate Explainable AI, start by defining your evaluation's purpose and imagining the results you need. This will guide your choice of methods. For example, to measure decision speed influenced by XAI, set up a realistic task using a prototype. You can then time users and observe where they struggle or seem uncertain.

For a well-rounded assessment, combine qualitative and quantitative methods:

- **Use Surveys**: When you have enough participants, surveys are great for collecting broad data on user opinions, preferences, and how well they think the explanations work. Use established surveys like those from Silva et al. [12] or Hoffman et al. [13] for reliability.
- **Conduct Interviews**: Go beyond surveys to get richer details. Use survey results as a starting point for deeper conversations.
- **Observe Users & Give Tasks**: Watch people use the XAI explanations in realistic situations. This helps identify usability problems that surveys or interviews might not reveal.

[12] Silva, A., Schrum, M., Hedlund-Botti, E., Gopalan, N., & Gombolay, M. (2023). Explainable artificial intelligence: Evaluating the objective and subjective impacts of xai on human-agent interaction. *International Journal of Human–Computer Interaction, 39* (7), 1390-1404. https://doi.org/10.1080/10447318.2022.2101698

To summarize, conducting user evaluations is essential for developing explainable AI solutions that align with human needs in process industries.

Methods Toolkit

Work Domain Analysis

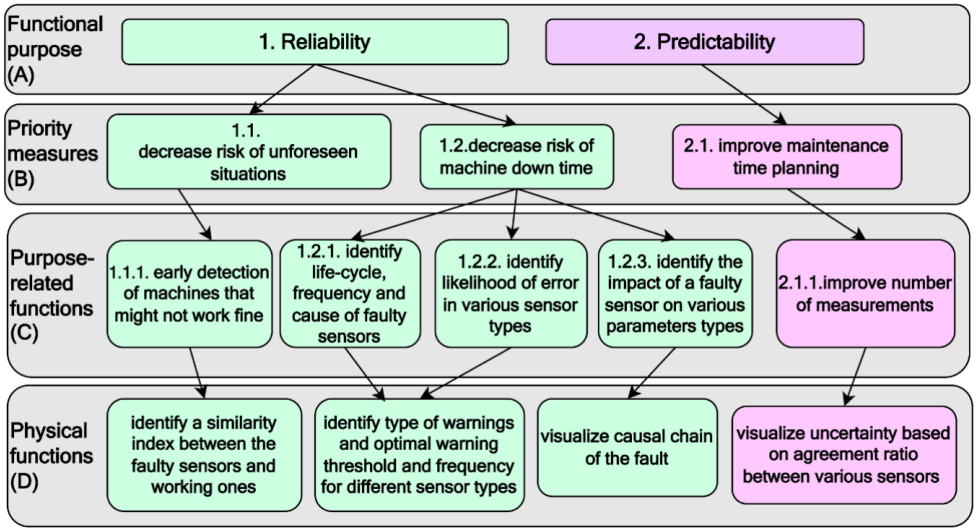


Figure 4 Work Domain Analysis for Vibration Monitoring use case (by Elmira Zohrevandi, Linköping University) [9]

Interview protocol used in the 1st phase of gathering user insights in EXPLAIN: main questions

- What is the process?
- Why is this process important?
- What are the risks?
- Who is/are the operators involved in the process? What are their goals? What educational background do they have?
- Are AI applications currently used?
- What type of support does the operator receive from these systems?
- What support could be improved, or would be welcomed?
- Are the systems always correct?
- Do you trust these systems?
- Do you encounter any problems with these systems? If so, what are common problems/issues?
- Are there system aspects that can be difficult to understand?
- What attitudes does the operator have towards the ability to provide feedback to the system?
- What risks could be associated with a feedback component?
- What attitudes are there around explainable system suggestions?
- What parts of the industrial process would benefit from system suggestions?

Explanation Satisfaction Scale (ESS) by Hoffman et al. [13]

1.From the explanation, I understand how the [software, algorithm, tool] works.
2.The explanation helped me understand why the [software, algorithm, tool] did what it did.
3.The explanation was easy to understand.
4.The explanation was informative.
5.The explanation was complete.
6.The explanation provided the right amount of detail.
7.The explanation was useful.
8.I am satisfied with the explanation I received.
9.Overall, the explanation was good.

For each item, users are typically asked to indicate their level of agreement on the 5-point scale. Researchers may also provide an optional "free response" opportunity for users to elaborate on their ratings. It's worth noting that researchers may choose to use a subset of these nine items depending on the specific goals of their study.

[13] Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). Metrics for explainable AI: Challenges and prospects. *arXiv preprint arXiv:1812.04608. https://doi.org/10.48550/arXiv.1812.04608*

Guidelines

6. How to Design and Develop Human-Centered Explanation and Feedback for Industrial AI Solutions

This chapter summarizes key learnings of designing explainer and feedback mechanisms for AI solutions for end users. In the project, we have investigated the needs and designed solutions for both operators in the process operations and data scientists who aim to improve the machine learnings models. However, more focuses were put to the solutions for operators. In the following, we will focus on the insights for XAI solution designs for operators in process industries.

Chapter Structure

6.1 The Users in Process Industries

6.2 Design XAI for Time Series Forecasting

6.3 Design XAI for Time Series Anomaly Detection

6.4 Design XAI for Image Analysis

6.5 XAI Insights and Recommendations: Differences Across Data Types

6.6 Key Insights for Feedback Mechanism

6.7 Conclusion: Explanations Need to Be Social

6.1 The Users in Process Industries

Given the complexity and critical nature of process industries, a range of specialized roles is essential to ensure smooth operations. Each role contributes unique expertise to manage, control, and optimize the process environment. In our project, we focused on the following roles related to specific use cases.

Data Type: Time Series Forecasing

Power Plant Dispatcher

Responsibilities
Predict energy demand and align plant operations accordingly.
Goals
- Match plant output to shifts in energy demand
- Understand current plant conditions and capabilities to operate at high capacity

Challenges
High cognitive demand, expertise required for handling conflicting data sources.

Control Room Operator

Responsibilities
Monitor and adjust plant operations based on forecasting and real-time data.
Goals
- Keep the process stable
- Achieve a quality that is above the goal threshold
- Be proactive with process adjustments.

Challenges
Declining data quality, automating fault detection.

Data Type: Time Series Anomaly Detection

Vibration Analyst

Responsibilities
Monitor vibrations in equipment to detect anomalies.
Goals
- Decrease risk of unforeseen situations
- Decrease risk of machine down time
- Improve maintenance time planning
- Improve operators understanding of potential incidents and challenges

Challenges
Identifying the right alarm settings, interpreting anomaly sources.

Power Plant Operator

Responsibilities
Monitor power demand and optimize energy generation.
Goals
- Meet the power demand of the current time block
- Recognize, understand and resolve anomalies and errors

Challenges
Understanding trigger alarms, high variability in sensor readings.

Data Type: Image Analysis

Inspection Operator

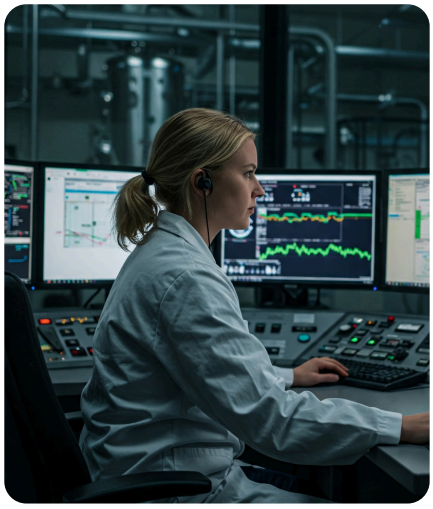
Responsibilities
Ensure high-quality inspections and defects detection in production.
Goals
- Ensure that the inspected device meet the required standards and specifications
- Optimize production and assembly of devices that align with customer expectations

Challenges
High cognitive demand, difficulty in automating defect assessment.

Simplified role descriptions that highlight key goals and challenges for specific use cases

6.2 Design XAI for Time Series Forecasting

Operators and process dispatchers working with time series data in industries like power generation, mineral processing, and pulp production share a common goal: optimizing complex, continuous processes. They balance numerous interrelated parameters to meet strict production targets, maintain quality, and ensure overall process stability.



Images generated by Google Gemini

Power plant Dispatcher

Data Type
Time Series Forecasting
Responsibilities
Predict energy demand and align plant operations accordingly.
Goals
- Match plant output to shifts in energy demand
- Understand current plant conditions and capabilities to operate at high capacity

Challenges
High cognitive demand, expertise required for handling conflicting data sources.

Control Room Operator

Data type
Time Series Forecasting
Responsibilities
Monitor and adjust plant operations based on forecasting and real-time data.
Goals
- Keep the process stable
- Achieve a quality that is above the goal threshold
- Be proactive with process adjustments.

Challenges
Declining data quality, automating fault detection.

User Challenges
The interplay of multiple variables makes it difficult to isolate the impact of any single factor affecting the processes, so errors or inconsistencies in forecasts can lead to significant disruptions. Domain expertise and tacit knowledge remain critical, meaning that any machine learning predictions should aim to augment rather than replace operators' mental models and hands-on experience. Processes often run continuously, creating an expectation of real-time or near-real-time predictions that can support immediate interventions and corrective measures. In some industries, such as the digester use case in pulp and paper, operators often face the complexity of the task imposed by the temporal aspects, for example, changes made to the process take several hours before their effects become observable.

The Need for XAI
Accurate and timely machine learning predictions could bring great value to these groups, as forecasts can support decision making related to operations, scheduling, or process adjustments. The reliability and transparency of predictive models at critical points can significantly influence operations, since large prediction errors or a failure to understand the model can disrupt production, reduce product quality, or result in financial losses. Thus, operators need explanations of why a model produces specific forecasts, particularly when outcomes may have a major operational or financial impact. The operators also benefit from contextual insights that help them understand the influence of different parameters, compare current readings to historical data, and detect outliers or unusual trends.

Key Insights

In this project, through three iterations, we have generated the following key insights about explainability in time series forecasting within industrial settings such as pulp and paper, mineral processing, and power plants.

Explainability as an interactive process
Explainability should be seen as an ongoing, interactive process. Operators often have specific questions about predictions, and explanations must evolve to address follow-up inquiries. This dynamic process ensures a deeper understanding of the model's predictions and the factors influencing them. User interface design plays a crucial role in making explanations interactive.

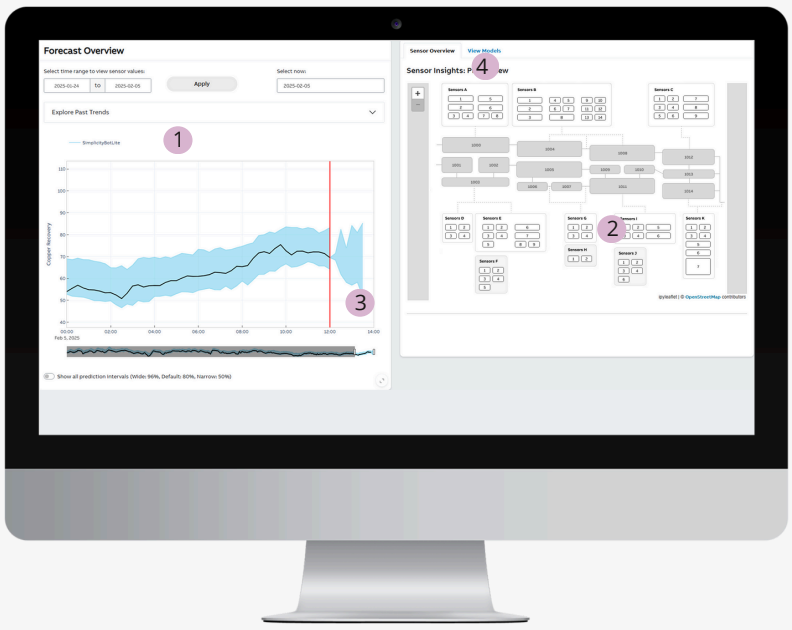
Integrating explainability into industrial workflows
Integrating explanations effectively into the workflow depends on factors such as the process interface, forecasting model, end-user preferences, and technical feasibility. Embedding explanations directly within process displays can allow users to relate AI insights to observed behaviors, enhancing comprehension. However, separating the explainer from the process interface can reduce confusion, although it may increase cognitive load by competing for the operator's attention.

Historical comparisons and counterfactual analysis
Allowing operators to compare current situations with historical scenarios provides valuable context, especially in non-time-critical situations. Counterfactual analysis, which lets operators explore how different actions might impact the process, supports scenario planning and risk assessment.

Prediction uncertainty is important
Uncertainty plays a key role in operator decision-making for complex industrial processes. Since simulations can't fully capture external disturbances or inherent process variability, adjustments prescribed by simulated control models cannot be fully trusted. Moreover, AI models often provide predictions without indicating the level of uncertainty, which can mislead operators. Forecasting systems should explicitly communicate uncertainty, allowing operators to evaluate the reliability of predictions and make more informed decisions.

Aligning with operators' mental models
Aligning AI systems with operators' mental models is challenging. Traditional methods like feature importance often rely on correlations, while operators focus on causal relationships between process parameters. This can create discrepancies between human reasoning and AI insights, opening the door for more interactive discussions that better align AI outputs with operator expertise.

Use Case Example



Dashboard design for the flotation process in mining industry. Developed by ABB Corporate Research Center in collaboration with Boliden, Sweden

1. Reference-based Explanation
How does this instance compare to a similar situation in the past?
2. Sensor plotting
What are the current sensor trends?
3. Uncertainty Band
How certain is the model prediction?
4. Model List
Which model provides the most accurate prediction in this situation?

Problem

Complex processes and delayed effects make predictions hard to trust without transparency—driving a strong need for explainable AI to support reliable, informed predictions.

User evaluation results

- Users' prior habits and mental models significantly shape how they interact with and interpret the prototype.
- **Uncertainty band:** Participants varied in grasping the concept of uncertainty bands.
- **Sensor plotting:** Operators found the sensor overview more useful than model lists.
- **Model list:** Users' appreciated the ability to explore different models and their associated predictions, providing flexibility to the system.
- **Reference-based explanations:** Seeing past successful strategies helps operators handle similar current situations. However, reference-based explanations in the prototype can be confusing as the historical trends are overlaid on current data.
- Some users believe ML systems can outperform them by incorporating more data sources and parameters.
- The model must demonstrate reliability and trustworthiness through consistent and accurate predictions.

6.3 Design XAI for Time Series Anomaly Detection

In industrial settings where complex machinery is continuously monitored and maintained - ranging from pulp and paper plants equipped with extensive vibration sensors to coal-fired power stations requiring round-the-clock oversight - operators and analysts share the goal of detecting and addressing potential issues before they escalate into costly failures or disruptions.

Images generated by Google Gemini

Vibration Analyst

Data Type
Time Series Anomaly Detection
Responsibilities
Monitor vibrations in equipment to detect anomalies.
Goals
- Decrease risk of unforeseen situations
- Decrease risk of machine down time
- Improve maintenance time planning
- Improve operators understanding of potential incidents and challenges

Challenges
Identifying the right alarm settings, interpreting anomaly sources.

Power Plant Operator

Data Type
Time Series Anomaly Detection
Responsibilities
Monitor power demand and optimize energy generation.
Goals
- Meet the power demand of the current time block
- Recognize, understand and resolve anomalies and errors

Challenges
Understanding trigger alarms, high variability in sensor readings.

User Challenges
User groups in this category monitor large volumes of data to identify anomalies, requiring a fair bit of experience to spot irregularities. Their work is challenged by the rapid increase in sensor data, the complexities of tuning alarm thresholds, the scarcity of labeled historical data for model training, and the inherent “black box” nature of many AI models, which can undermine trust if operators do not understand why certain anomalies are flagged. These challenges occur in control room environments where users manage multiple screens to discover irregularities and make decisions.

The Need for XAI
To make it easier and faster to detect anomalies, these users need reliable, real-time, and explainable AI models that help them identify and manage critical events efficiently, without overwhelming them with unnecessary or unclear alerts. Though many are open to AI systems in their workplace, they emphasize the importance of retaining control over final decisions. They prefer solutions that integrate seamlessly into existing workflows, with clear explanations available when an anomalous situation arises. This ensures that domain experts can quickly verify whether an alarm is meaningful, while less experienced operators benefit from interpretable guidance to take appropriate actions. In this context, the ultimate objective is to minimize unplanned downtime and improve overall operational safety and efficiency, aided by AI-based anomaly detection that is transparent, explainable, and adaptable to evolving data streams.

Key Insights

In this project, through three iterations, we have generated the following key insights about explainability in time series anomaly detection within industrial settings such as pulp and paper and power plants.

Explainability as a multi-faceted approach
Explainability should not rely on a single technique but rather integrate multiple complementary explanation methods. Techniques such as anomaly scores, heatmaps, frequency attribution, and reference-based explanations provide operators with a comprehensive understanding of anomalies, aiding in decision-making in different contexts.

Meta explainers are important for industrial applications
The solution should enable iterative refinement of explanations—starting with basic outputs like attribution maps or classifier results and evolving into more intuitive formats such as natural language summaries or contextualized visualizations. By using one explainer's output to enhance another, explanations become clearer and more relevant, reducing cognitive load and supporting quicker, more confident decision-making. This is especially valuable in industrial contexts, where operators need varying levels of explanation depending on expertise and situation.

Context is key and reference-based explanation
Interpretability of found anomalies would be enhanced by providing additional relevant data points or metadata. Contextualization supports operators in making informed decisions by placing anomalies within a broader operational or environmental perspective. This includes utilizing past reference-based explanations which allows operators to compare current anomalies with historical data or exemplary cases. This comparison contextualizes anomalies, facilitating quicker recognition of patterns and more effective decision-making.

Prediction uncertainty is important
Clearly communicating prediction uncertainty through confidence intervals or uncertainty metrics is crucial for enhancing operator trust and decision-making quality. Transparent uncertainty measures empower operators to distinguish critical anomalies from less significant fluctuations effectively.

Use Case Example

Dashboard design for anomaly detection use case in power plant. Developed by ABB in collaboration with LEAG, Germany

1. Reference-based Explanation
Has it happend before?
2. Natural Language Explanation
Can the AI-predictions be rephrased in a sentence?
3. Feature Importance
Where is the anomaly coming from?
4. Counterfactual Explanation
How should the signal trend have looked like?
5. Contextualization
Are there additional contextual information providing further insight?

Problem

Anomaly detection often lacks of explainability of the AI black box models, and the lack of known historic anomalies (labeled data) when building these models.

User evaluation results [14]

- The dashboard was seen as a valuable tool for identifying and narrowing down anomalies.
- Operators emphasized that AI should only alert, not act, leaving final judgment to them. They preferred the dashboard to be available on demand rather than always visible.
- **Reference-based explanations** were considered as very helpful, especially when linked with logbook entries to suggest solutions.
- **Natural language explanations** were added following the evaluation, and are believed to be valuable for offering quick and accessible insights into anomalies.
- **Feature importance** was rated as the most helpful feature. During troubleshooting it would help guide the field operator to find the location of the issue and rectify it.
- **Counterfactual explanations** were mostly considered as very helpful. A comparison between actual-vs-normal values does not exist in their current systems today, meaning that a lot of manual work is required to make a similar comparison.
- **Contextualization** was seen as helpful, as it helps quickly identify which alarms are related to an anomaly, saving time and improving decision-making speed, compared to manually searching through control system alarms.

[14] Dix, M., Koltermann, J., Mieck, S., Pastler, B., & Kloepper, B. (2024). XAI for anomaly analysis by power plant operators-a case and user study. *In ML4CPS–Machine Learning for Cyber-Physical Systems.* UB HSU.

6.4 Design XAI for Image Analysis

In electronics manufacturing, guaranteeing high-quality products involves multiple inspection steps following PCB (Printed Circuit Board) assembly or final product assembly. The inspection types include checks for presence, alignment, or damage.

Inspection Operator

Data Type
Image Analysis
Responsibilities
Ensure high-quality inspections and defects detection in production.
Goals

- Ensure that the inspected device meet the required standards and specifications
- Optimize production and assembly of devices that align with customer expectations

Challenges
High cognitive demand, difficulty in automating defect assessment.

Images generated by Google Gemini

User Challenges
Visual inspection operators at the End-of-Line (EOL) stage are responsible for determining whether a product passes or fails based on defined criteria and their own judgment. Working at stations equipped with inspection tools and digital systems, they perform manual checks guided by inspection forms and visual references, considering factors like physical condition, assembly, and customer-specific requirements. For smaller or less complex products, machine learning–based computer vision assists by flagging potential defects, but operators retain final decision authority. They also report faults via internal systems, contributing to quality improvements in manufacturing and assembly. Key challenges include balancing standardized checks with tacit knowledge, handling varied product complexity under time constraints, and verifying ML outputs while maintaining accountability for quality decisions.

The need for XAI
There is a need for models that clearly explain why a product is flagged. Operators need insights that directly relate model outputs to established quality standards—such as which specific inspection criterion was not met, what part of the product caused the failure, and where this is visible in the camera image. Additionally, understanding the training data used by the model, including who labeled it, who developed the model, and when it was last updated, is important for transparency and accountability. Familiar with image-based feedback, operators expect explanations to visually highlight the defective areas tied to model decisions. Enhancing these capabilities can improve decision speed, trust, and overall alignment between human judgment and AI support in the inspection process.

Key Insights

In this project, through three iterations, we have generated the following key insights about explainability in image analysis within industrial settings such as PCB production and inspection.

Explainability as a multi-faceted approach
Combining different methods could help users understand better about the current situations, the machine learning output and sometimes also what actions may take. The combination will also ensure usability across different user expertise levels.

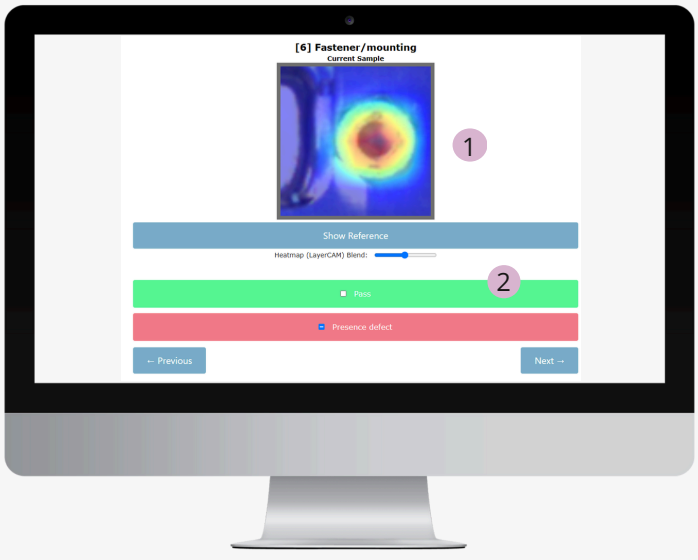
Customizability to enhance practical adoption
The system should be adjustable based on user expertise, allowing operators and engineers, to access explanations at different granularity levels. Interactive tools that enable users to tweak the explanation scope, such as selecting different layers of heatmaps or filtering feature importance, improve usability and adaptability in diverse industrial settings.

Industry specific considerations for image processing explainability
Different industries require tailored explainability approaches. Explainability must align with regulatory compliance, ensuring decisions can be audited and justified.

Explanation management
Not all explanations are useful for all users. Filtering mechanisms should remove redundant, overly complex, or irrelevant explanations while preserving interpretability. This can be especially important in feature attribution methods, where low-impact features should be filtered out to focus on the most critical contributors to a model's decision. Implementation of adaptive explainability mechanism might improve the dynamic adjustment of explanation based on users requirements or expertise.

Continuous monitoring of explanation effectiveness beside model performance
Involving domain expertise is important for continuous improvement of explanations and models. Users should be able to give feedback on prediction accuracy, though this depends on their expertise. Further, metrices should track explanation consistency, user trust, and error rates to ensure that explanations remain reliable. Logging user interactions and feedback can help identifying gaps where the model fails to provide meaningful justifications, supporting long-term improvement.

Use Case Example



Dashboard design for visual inspection use case in electronic manufacturing. Developed by Prodrive in collaboration with Marantz-Electronics (MEK), the Netherlands

1. Heat map
Which pixels were most important for this prediction?

2. Reference-based explanation
What does a compliant component look like?

Problem

To ensure high-quality products, electronics manufacturing involves multiple inspections throughout the production process to check for presence, alignment, and damage. Many of these are manual or use non-deep-learning algorithms, leading to missed defects, high false positives, and poor defect traceability.

User evaluation results

- Multiple methods were appreciated by the users. A **heatmap** highlights the defect, while **a reference-based explanation** compares it to past similar cases, offering a comprehensive understanding.
- The **reference explainer** is considered very helpful to indicate how a compliant component looks like.
- Users suggested that unclear or incorrect explanations could signal model misinterpretation of certain images. Incorporating these images into the training set could enhance model robustness, linking explanation review with iterative retraining.
- Users valued the use of anomaly detection techniques. Visualizing samples in a 2D distribution plot was seen as helpful for identifying out-of-distribution data and understanding model behavior.
- Users expressed a desire for a more interactive approach. This is important given that they will inspect a large amount of images per day.
- Some users proposed adding explanations to a data management dashboard. Displaying heatmaps alongside original images and model predictions could enhance transparency and user insight during analysis.

6.5 XAI Insights and Recommendations - Differences Across Data Types and Use Cases

While XAI principles share many similarities across time series forecasting, time series anomaly detection, and image analysis, key differences arise due to the nature of the data and domain-specific challenges. The following table summarizes these differences.

Table 1 Different Characteristics among Data Types

	Time Series Forecasting	Time Series Anomaly Detection	Image Analysis
Nature of Data	Continuous, sequential patterns over time.	Event-based, detecting irregular deviations in time series.	Spatial data, often requiring object detection and pattern recognition.
Key Explainability Challenge	Aligning AI outputs with operators' mental models and managing forecasting uncertainty.	Addressing false positives/ false negatives and ensuring anomalies are relevant.	Ensuring accurate defect classification and explaining visual reasoning.
Role of Context	Historical data comparisons and counterfactuals help operators understand predictions.	Reference-based explanations help validate if an anomaly is significant.	Visual reference comparisons are essential for defect identification and classification.
Human Interaction	Support interactive exploration of what-if scenarios and historical trends.	Explanations should quickly clarify anomaly significance to avoid alarm fatigue.	Users need adjustable explanations to match expertise levels.

The table summarizes the differences among three data analysis types: Time Series Forecasting, Time Series Anomaly Detection, and Image Analysis. It compares them based on the nature of the data, key explainability challenges, the role of context, and human interaction aspects for each type.

Table 2 Design Implications for Various Data Types

	Time Series Forecasting	Time Series Anomaly Detection	Image Analysis
Explanation Methods	Hybrid models using SHAP, feature attribution, temporal trends, and counterfactual explanations	Anomaly scores, heatmaps, frequency-based attributions, and reference-based methods.	Heatmaps, hierarchical explanations, and contrastive visual comparisons.
Uncertainty Handling	Clearly communicate forecast confidence levels and probability distributions.	Provide confidence intervals to separate critical from non-critical anomalies.	Implement uncertainty-aware visual indicators for AI-detected defects.
Workflow Integration	Embed explanations within process control interfaces without overloading users.	Ensure real-time, non-intrusive explanations that do not overwhelm operators with excessive alerts.	Allow users to zoom into specific features or switch between explanation layers.
User Adaptability	Adapt explanations for operators making long-term decisions.	Adapt explanations for operators monitoring real-time anomalies.	Provide multi-resolution hierarchical explanations for engineers and operators.
Human-AI Interaction	Enable interactive scenario exploration.	Allow operators to refine anomaly detection thresholds based on feedback loops.	Provide interactive image analysis tools to refine defect detection accuracy.

This table provides design implications for these three different data analysis types. It outlines specific approaches for each type across several key areas, including explanation methods, how to handle uncertainty, workflow integration strategies, ways to ensure user adaptability, and methods for human-AI interaction.

Despite differences in the nature of time series forecasting, time series anomaly detection, and image analysis, there are common insights in Explainable AI for all three types of use cases. One key overarching principle is that Human-AI collaboration is essential. Users should retain control over AI decisions, with explanations designed to support—not replace—human expertise. The foundation of this is that operators should have adequate understanding of AI models or outcomes, to be able to effectively control the system.

Table 3 Insights and Design Implications crossing Data Types

Key Insight	Description	Design Implications	Description
Explainability as a Multi-Faceted Process	Hybrid models using SHAP, feature attribution, temporal trends, and counterfactual explanations	**Use Hybrid Explanation Approaches**	Combine techniques like SHAP, feature attribution, heatmaps, and counterfactuals etc. for more comprehensive insights.
Context is Crucial	Explanations should relate AI insights to historical data, reference cases, or past trends to improve interpretability.	**Develop Reference-Based and Contextual Explanations**	Anomalies, defects, or predictions should be compared with past cases to help users understand deviations and trends.
Prediction Uncertainty Matters	AI systems must communicate confidence levels and uncertainty to help users assess prediction reliability and trust AI-driven decisions.	**Implement Uncertainty-Aware Explanations**	AI outputs should indicate reliability through confidence scores, probability distributions, or uncertainty metrics.
Workflow Integration is Key	Explanations should seamlessly fit into industrial workflows, ensuring clarity without causing cognitive overload.	**Ensure Real-Time Interpretability**	In time-sensitive industrial settings, explanations should be available instantly to aid fast decision-making.
		Develop industry-specific explanation metrics	Customize explainability metrics to match domain-specific evaluation needs.
		Design for Actionable Explanations	Causal information can help operators—but only if the process is well-modeled and the interface presents it clearly and at the right time.
Adaptability for Different Users	Explanations should be customizable based on user expertise, providing different levels of detail for operators, engineers, and analysts.	**Provide Multi-Layered and Customizable Explanations**	Explanations should be structured progressively, allowing different users to access insights at various levels of granularity.
		Incorporate Interactive Explainability	Users should be able to interact with explanations by adjusting inputs, choosing detail levels, or testing counterfactuals. Prompting the explainer further lets users deepen their understanding of the AI model's decisions.

This table outlines key insights and their design implications for building effective explainable AI systems. It emphasizes that explainability is complex and requires combining multiple techniques. Contextual information, such as historical data and past trends, is crucial for understanding AI explanations. The table also highlights the importance of communicating prediction uncertainty to build user trust. Seamless integration of explanations into existing workflows is essential for real-time interpretability. Finally, it stresses the need to adapt explanations to different users' expertise levels and to allow for interactive exploration of the explanations.

User Challenges and Explanation Methods Mapping

As industries increasingly integrate Artificial Intelligence (AI) into their operations, the need for Explainable AI (XAI) becomes crucial. Our industrial partners have shared valuable insights on how XAI can be effectively developed, implemented, and leveraged in real-world settings. This chapter highlights the key learnings derived from industry practitioners, covering benefits, challenges, and best practices for successful XAI adoption.

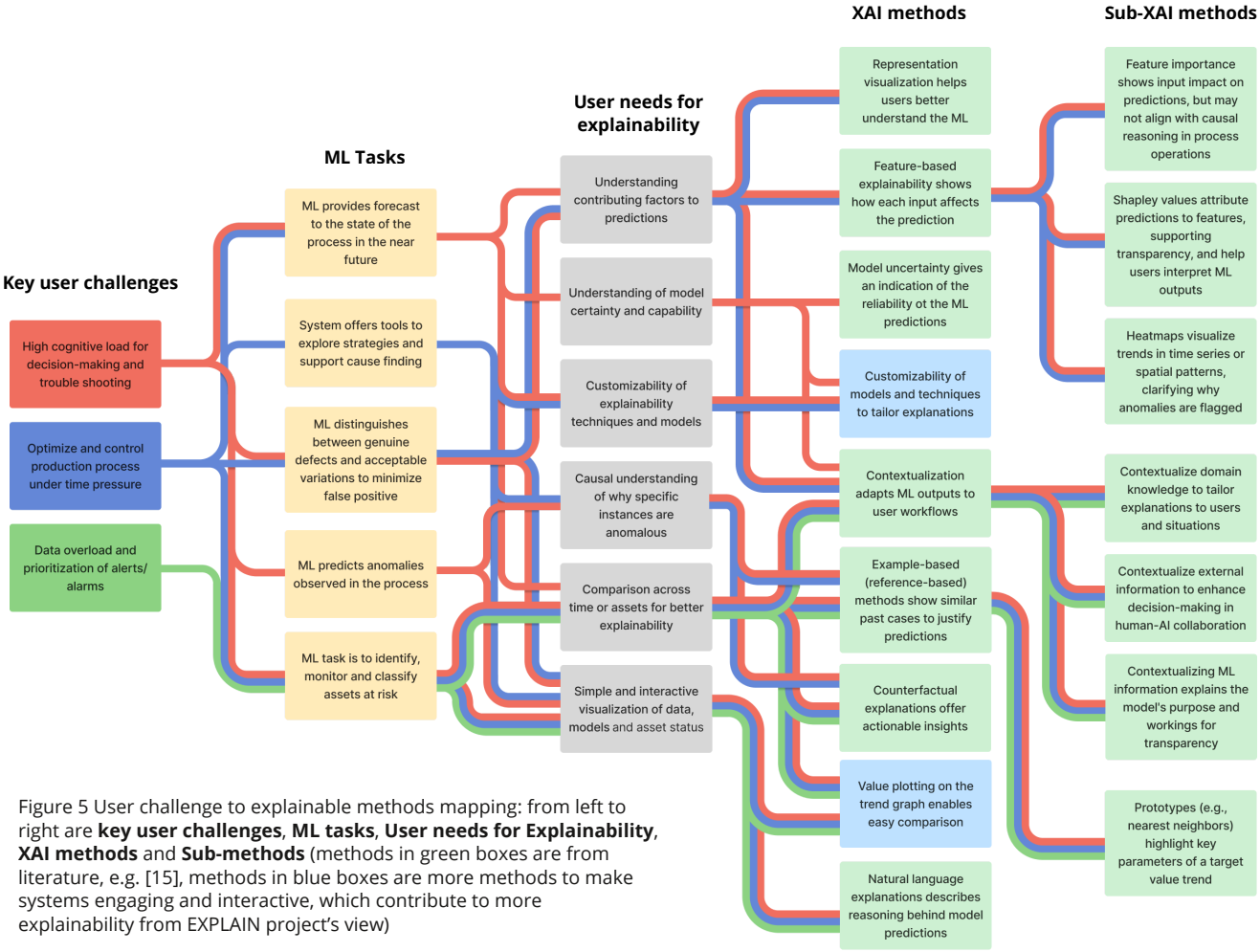


Figure 5 User challenge to explainable methods mapping: from left to right are **key user challenges**, **ML tasks**, **User needs for Explainability**, **XAI methods** and **Sub-methods** (methods in green boxes are from literature, e.g. [15], methods in blue boxes are more methods to make systems engaging and interactive, which contribute to more explainability from EXPLAIN project's view)

This diagram provides a comprehensive mapping between key user challenges in industrial settings and relevant XAI methods. It starts by identifying user challenges such as high cognitive load, the need to optimize processes under time pressure, and data overload. These challenges were summarized from our user research insights. It then links these challenges to specific ML tasks, like forecasting, offering support strategies, minimizing false positives, predicting anomalies, and classifying assets. Crucially, the diagram connects these ML tasks to specific user needs for explainability, such as understanding contributing factors, model certainty, the ability to customize explanations, understanding the cause of anomalies, and the need for simple visualizations and comparisons. Finally, it maps these user needs to both general and sub-explainable methods. General methods include representation visualization, feature-based explanations, model uncertainty indicators, customizable models, example-based reasoning, natural language explanations, counterfactuals, workflow contextualization, and value plotting. Sub-explainable methods offer more specific techniques like feature importance, Shapley values, heatmaps, prototypes, and various forms of contextualization (domain knowledge, external information, and ML information itself).

This diagram demonstrates a structured, user-centric approach to XAI. It moves beyond simply listing XAI techniques, instead directly links them to real-world user challenges and the specific informational needs arising from their interaction with ML systems. This mapping ensures that the chosen explainability methods are relevant, actionable, and truly address the users' pain points in their work context. The emphasis on contextualization as a key element across different stages of the explanation process also highlights a significant aspect of making AI truly understandable and useful for human users in complex industrial environments.

[15] Lai, V., Chen, C., Smith-Renner, A., Liao, Q. V., & Tan, C. (2023). Towards a science of human-AI decision making: An overview of design space in empirical human-subject studies. *In Proceedings of the 2023 ACM conference on fairness, accountability, and transparency* (pp. 1369-1385). https://doi.org/10.1145/3593013.3594087

Explainability and Industrial Problems Mapping

The following diagram shows the explanation methods which have been experiemented and applied for the use cases within EXPLAIN project.

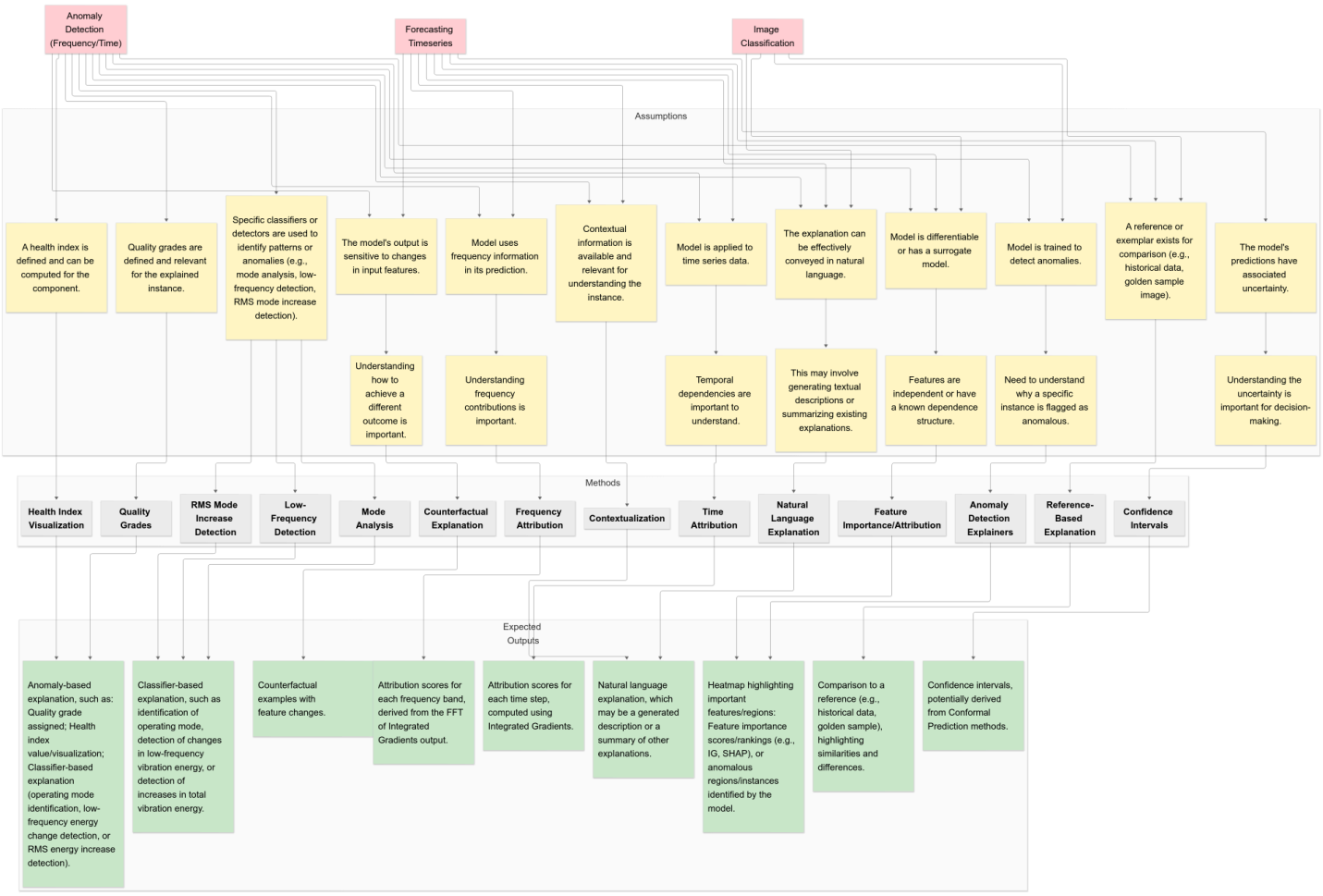


Figure 6 Explainability and Industrial problems mapping: from top to bottom levels: Industrial problems or data types, technical assumptions, methods, expected outputs

This diagram illustrates the mapping between different data types and various Explainable AI (XAI) methods. It showcases a clear and structured approach to connecting specific data types with appropriate XAI methods based on the inherent characteristics and explainability requirements of each data type.

For **Anomaly Detection**, the focus is on understanding why a particular instance is flagged, the importance of feature combinations, and the significance of attribute frequencies. Corresponding XAI methods include health index visualization, quality metric display, RMS mode detection, low frequency detection, mode analysis, counterfactual explanations, and frequency attribution.

For **Forecasting**, where understanding temporal aspects is crucial, the diagram highlights the need for explanations involving textual descriptions, comparisons across time instances, and the identification of features with temporal dependencies.

For **Image Classification**, the primary explainability need is to understand which parts of the image influenced the decision. Reference-based explanations and confidence scores are highlighted as relevant methods.

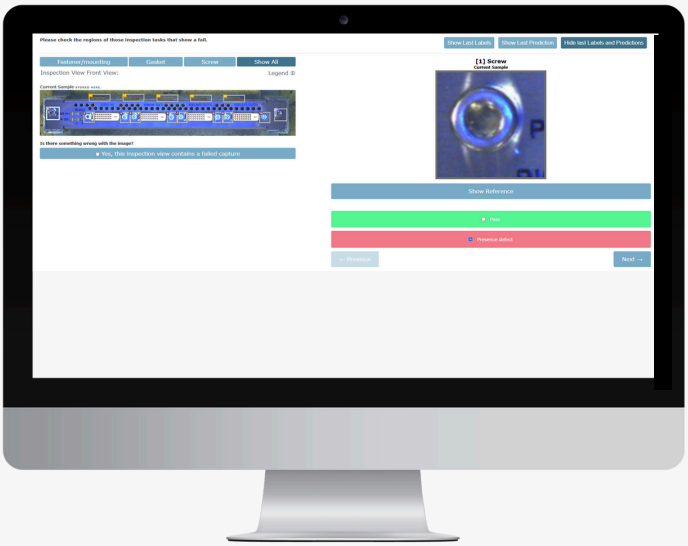
This diagram provides a tailored framework that considers the unique aspects of different data modalities. It also offers practical guidance for selecting the most effective explanation techniques, ultimately leading to more meaningful and actionable insights for users working with diverse data.

6.6 Key Insights for Feedback Mechanism

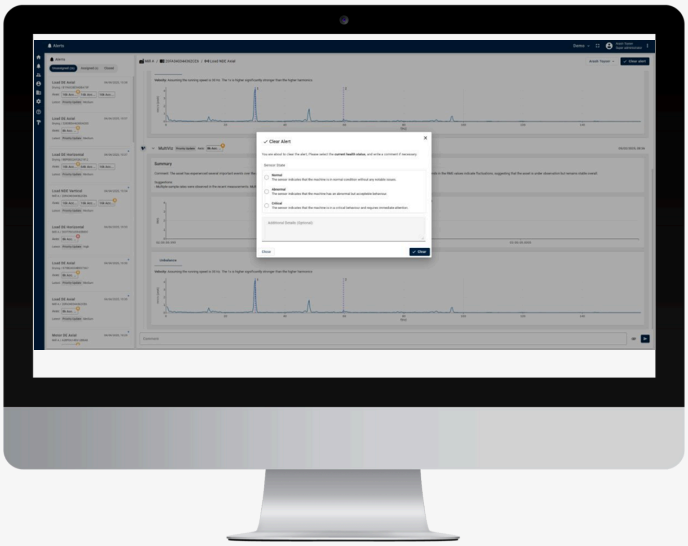
In this project, we have explored different ways of giving feedback, and here is a summary of key insights for designing effective feedback mechanisms to refine and train ML systems:

- **Explainability as a prerequisite for quality Feedback:** The system must provide sufficient explainability for operators to give meaningful feedback.
- **Offer diverse feedback channels:** Providing multiple ways for users to submit feedback caters to different preferences and working styles. This could include options like direct annotation on outputs, structured forms, free-text fields, or even voice input, depending on the application.
- **Enable user understanding:** Invest in making the ML model's outputs or underlying logic understandable to the users who provide feedback. This could involve visualizations, explanations of predictions, or access to relevant data points. Informed feedback is significantly more valuable and encourages active participation.
- **Empower all stakeholders:** Ensure that feedback can be provided by all relevant users, including both operational staff and model development teams. Different perspectives can highlight different types of issues and improvement opportunities.
- **Experience-based feedback control:** Sometimes the domain expertise is crucial, then the feedback should be limited to qualified users. The system should support experienced users in providing accurate input.
- **Prioritize user convenience:** The feedback mechanism should be seamlessly integrated into the user's workflow and require minimal effort. Direct feedback options within the application interface, such as on a live dashboard, are crucial for reducing friction and increasing the likelihood of feedback.
- **Implement robust feedback governance:** Bias exists in feedback. Operators may label similar situations differently based on experience. Not all feedback should be weighted equally—experience level should be considered. It is important to develop clear processes for managing and utilizing the feedback received. This includes strategies for filtering out noise, identifying and mitigating potential biases in human feedback, and prioritizing feedback based on its impact and reliability. Techniques like aggregating feedback, assigning confidence scores, or having expert review processes can be beneficial.
- **Demonstrate the value of feedback:** Showing users how their feedback contributes to improvements in the system is a powerful motivator. This could involve communicating updates based on feedback, highlighting the impact of specific feedback instances, or visualizing the overall improvement in model performance resulting from user input. This transparency fosters trust and encourages continued engagement with the feedback process.

Use Case Examples



Labeling in the image classification use case
Feedback designed and developed by Prodrive in collaboration with Marantz-Electronics (MEK), the Netherlands



Clearing alerts for vibration analysis use case
A feedback interface designed and developed by Viking Analytics

Feedback methods and Industrial Problems Mapping

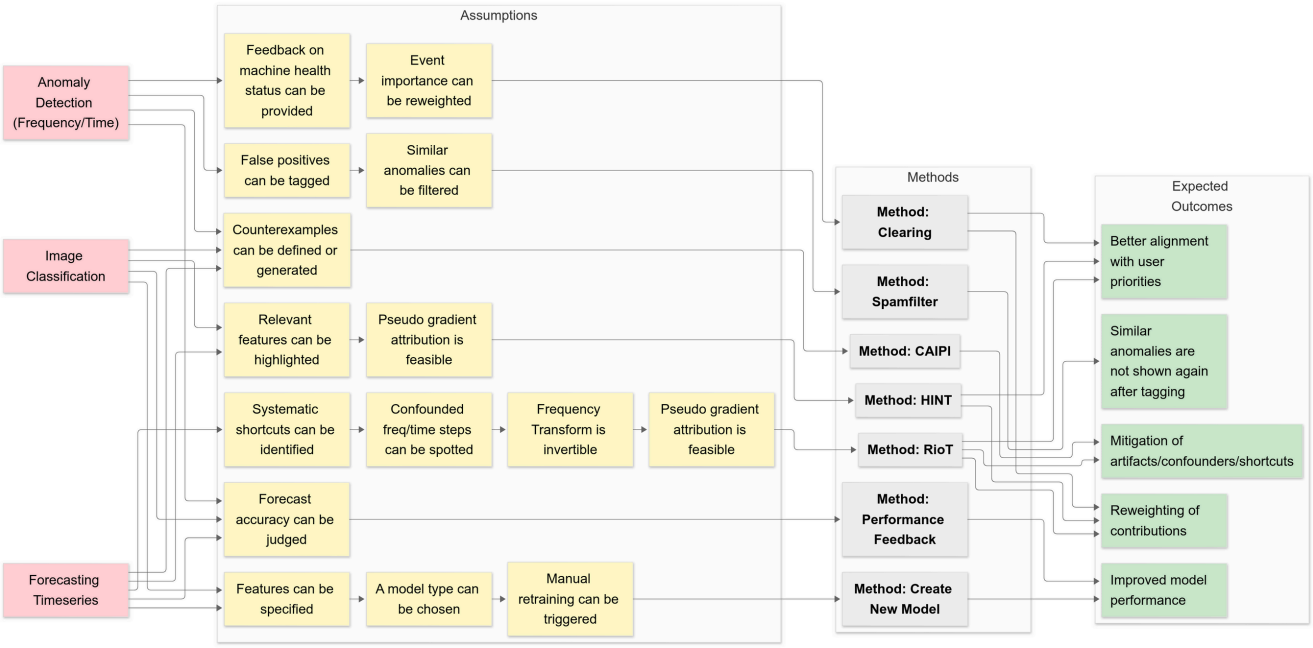


Figure 7 Feedback and Industrial problems mapping: from top to bottom levels: Industrial problems or data types, technical assumptions, methods, expected outputs

The diagram outlines a process where feedback is collected and utilized to refine machine learning models. It starts with different ML tasks and highlights several key assumptions or capabilities that enable effective feedback. These include providing feedback on machine health, tagging false positives, defining counterexamples, highlighting relevant features, identifying systematic shortcuts, and judging forecast accuracy. Additionally, it assumes the ability to reweight event importance, filter similar anomalies, utilize pseudo-gradient attribution, spot confounded frequency/time steps, invert frequency transforms, specify features, choose model types, and trigger manual retraining.

Based on these assumptions, various methods are employed to incorporate the feedback. These methods include Clearing, Spamfilter, CAIPI [16], HINT [17], RioT [18], Performance Feedback, and creating new models. The ultimate goal of the feedback loop is to achieve several expected outcomes, such as better alignment with user profiles, preventing the recurrence of similar anomalies, mitigating artifacts, confounders, and shortcuts, reweighting contributions, and ultimately improving overall model performance.

This diagram offers a holistic and integrated framework to feedback across diverse machine learning tasks. Instead of treating feedback as a separate process for each task, this framework proposes a unified system that leverages various types of feedback and applies a range of methods to address different model deficiencies.

It suggests a generalizable approach to model improvement through feedback. It incorporates various forms of feedback, from simple tagging and performance metrics to more sophisticated methods like defining counterexamples and highlighting relevant features. It links specific feedback capabilities to particular methods, indicating a thoughtful approach to addressing different types of model issues. The expected outcomes clearly target common problems in machine learning, such as misalignment with user needs, recurrence of errors, and the presence of undesirable artifacts or shortcuts.

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

[17] Selvaraju, R. R., Lee, S., Shen, Y., Jin, H., Ghosh, S., Heck, L., ... & Parikh, D. (2019). Taking a hint: Leveraging explanations to make vision and language models more grounded. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 2591-2600).

[18] Kraus, M., Steinmann, D., Wüst, A., Kokozinski, A., & Kersting, K. (2024). Right on time: Revising time series models by constraining their explanations. arXiv preprint arXiv:2402.12921.

6.7 Conclusion: Explanations Need to Be Social

The EXPLAIN project contributed to human-centered XAI research and development for process industries. The learnings related to the design and development are in line with the trend of seeing explanations more social than technical components.

The technical mechanisms explain the inner workings of complex machine learning models. These efforts are primarily aimed at providing insights into how a model arrived at a particular prediction, often employing techniques like feature importance scores, gradient-based saliency maps, or rule extraction. While these methods offered valuable information about the model's internal logic, they sometimes do not align with end users' mental models, making it hard for them to understand without training. End users, the operators in process industries in our case, are often less interested in the intricate details of the model's architecture and more concerned with understanding why a decision was made in a way that aligns with their understanding of the process and their goals. They need explanations that are intuitive, relatable, and provide a basis for trust and informed decision-making. The project results generally support that explanation is fundamentally a social act.

Explanations involves a dialogue (implicit or explicit) between the explainer (the AI system or the XAI method) and the explainee (the human user). Effective explanation requires considering the audience, their prior knowledge, their goals, and the specific context in which the explanation is being sought. Social explanations, therefore, move beyond merely revealing internal model parameters or mathematical relationships. Instead, they should aim to provide answers to questions like: "Why did the model make this decision for me?", "What would need to change for the model to make a different decision?", or "How does this decision align with my understanding of the situation?".

We addressed the social explanations in two ways; firstly, we have applied human-centered evaluation of explanations. Instead of solely relying on technical metrics of explanation fidelity, we have focused on evaluating how well explanations are understood, how much they increase user trust, and whether they lead to better decision-making. Secondly, there was a broad exploration of explanation formats that are more natural and intuitive for humans, such as visualizations and interactions, searching for similar cases in history, natural language explanations, counterfactual explanations etc..

The transition from technical explainer mechanisms to social explanations in XAI represents a move towards building AI systems that are not only accurate but also transparent and trustworthy in a way that resonates with human understanding and social norms. This shift is crucial for the widespread adoption and responsible deployment of AI in real-world applications, fostering a collaborative relationship between humans and intelligent machines.

7. Human-Centered MLOps for Trustworthy Industrial AI

In industrial applications, companies face several challenges, such as the integration of legacy systems, real-time processing within milliseconds, the quality and availability of sensor data, and the resilience of ML methods [19]. In this context, interpreting ML model inferences, integrating them into existing IT infrastructures, and incorporating expert knowledge during development and operations remain major challenges.

[19] Faubel, L., Woudsma, T., Kloepper, B., Eichelberger, H., Buelow, F., Schmid, K., ... & Bång, M. (2024). MLOps for cyber-physical production systems: challenges and solutions. *IEEE software.* https://doi: 10.1109/MS.2024.3441101

What is MLOps:
"MLOps (Machine Learning Operations) is a paradigm, including aspects like best practices, sets of concepts, as well as a development culture when it comes to the end-to-end conceptualization, implementation, monitoring, deployment, and scalability of machine learning products. Most of all, it is an engineering practice that leverages three contributing disciplines: machine learning, software engineering (especially DevOps), and data engineering" [20].

[20] Kreuzberger, D., Kühl, N., & Hirschl, S. (2023). Machine learning operations (mlops): Overview, definition, and architecture. *IEEE access, 11*, 31866-31879. https://doi: 10.1109/ACCESS.2023.3262138.

MLOps Processes
MLOps processes are particularly important for running and maintaining ML models alongside explainable AI (XAI) and feedback mechanisms. These processes include model training, data preparation, validation, deployment, and monitoring. The MLOps life cycle proposed in the research project EXPLAIN, visualized in Figure 8, includes and extends these components by distinguishing between the initial development phase (blue circle) and the production phase (green).

This life cycle significantly advances traditional MLOps approaches by incorporating explainability and feedback mechanisms (orange), and by actively involving domain experts and operators throughout the process. First, users and engineers are provided with tools to better understand why and how AI systems make predictions. They can also offer input on issues or malfunctions, thereby improving the trustworthiness, reliability, and resilience of the models. Furthermore, this life cycle enables stakeholders to participate in the design and development of AI solutions, ensuring that models are better tailored to their specific needs.

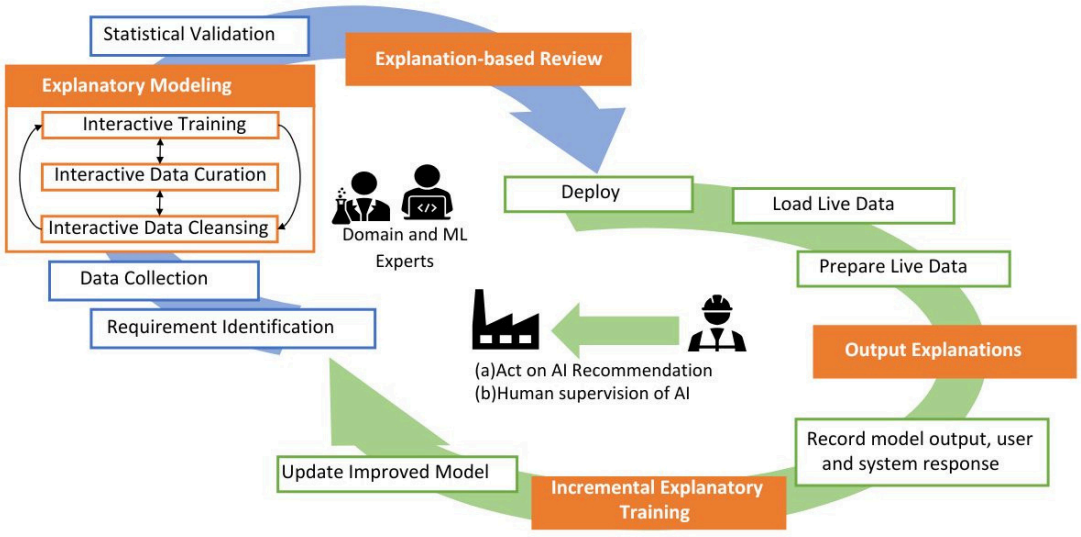
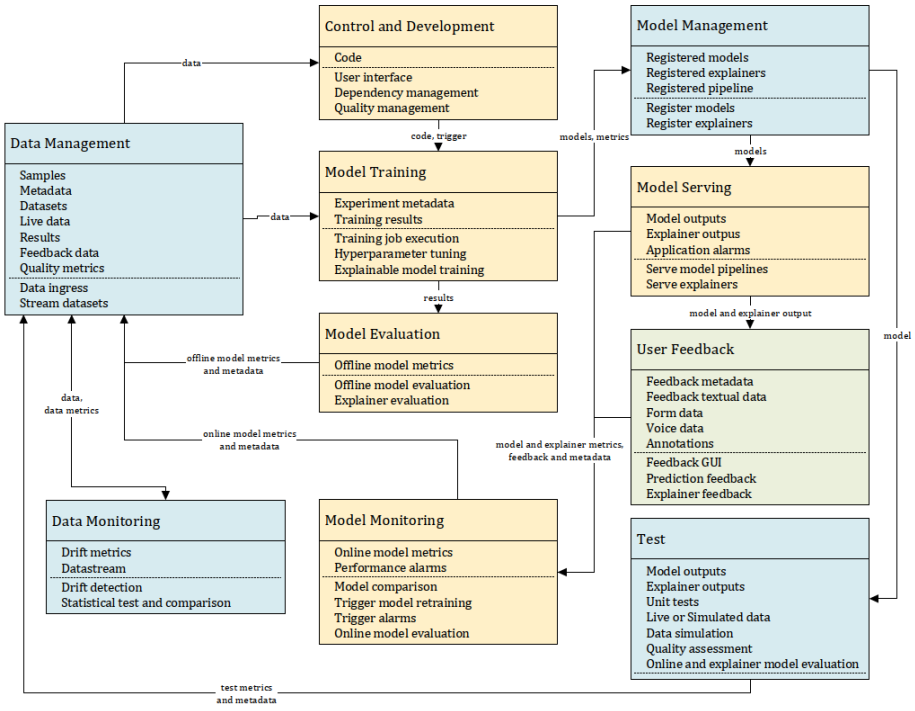


Figure 8 The MLOps process life-cycle proposed in the research project EXPLAIN



This architecture enables the integration of explainers, feedback, and interactive machine learning. It is implemented by several companies and is described as an example implementation in [21] Gitlab https://www.uni-hildesheim.de/gitlab/explain/mlops-environment-for-explanatory-ml-models.

Figure 9 The architecture of an MLOps system with explainer and feedback components proposed in the project EXPLAIN

Software Architecture

The architecture of an MLOps system with XAI and feedback components typically comprises several layers. Figure 9 shows the architecture used in the project EXPLAIN with the major components described below. The data management processes and stores large amounts of data, including feedback. Data management aims to detect drift and changes in the data as a potential trigger for retraining of the ML model. Model training and evaluation use scalable frameworks and integrated XAI methods. The model management registers the models, explainers, and their performance. Model Monitoring and serving are supposed to provide and observe models and explainers in the real world using container technologies. Finally, feedback integration mechanisms are used to capture feedback for maintenance and continuous improvement of the models.

Individual XAI methods can have a very high workload, requiring pre-processing and training. The MLOps infrastructure needs to take that into account. A solution is to run XAI as separate services. This requires scalable solutions and can be handled similar to ML algorithms.

Bridging Complexity: Tooling and Challenges in Industrial MLOps for Explainable AI

Developing MLOps systems remains complex, time-consuming, and error-prone, often relying on a mix of existing tools. Many companies adopt cloud platforms like AWS, Azure, or GCP for simplicity, where select XAI methods are integrated. However, user interaction and feedback automation in XAI are still lacking. Due to security, regulations, and cost, companies are reluctant to go into the cloud and develop their own in-house MLOps systems.

MLOps tools cover areas like data analysis, CI/CD, automation, testing, programming, versioning, orchestration, deployment, and monitoring. Tool choice depends heavily on project needs and infrastructure, requiring careful planning. Utility trees and tool catalogues can aid selection. While many tools support explainability, feedback mechanisms often need to be extended with self-developed feedback solutions.

In industry, ML deployment must handle legacy systems, real-time demands, and data quality. The EXPLAIN project shows that embedding explainability and feedback into MLOps improves transparency, trust, and user engagement. Though current tools address many needs, integrating human-centered design and technical reliability is essential for trustworthy AI.

8. Turning Concepts to Real Deployment in Industries: Business Perspectives

This chapter highlights the key learnings derived from industry practitioners, covering benefits, challenges, and best practices for successful XAI adoption in process industries.

Benefits

We have identified clear benefits of XAI solutions for process industries:

Enhancing Adoption and Utilization
XAI significantly improves the likelihood of AI adoption by making AI-driven recommendations more transparent and understandable. This fosters trust among operators, engineers, and decision-makers, encouraging wider use of AI technologies.

Improving Decision-Making
Industries dealing with complex processes, such as energy price forecasting or industrial optimization, benefit from XAI's ability to provide clear, interpretable insights. Operators can better assess AI-generated outputs, leading to more informed decisions.

Strengthening Interdisciplinary Collaboration
XAI initiatives bring together experts from various domains—AI, process engineering, human factors, and IT—leading to holistic solutions that cater to diverse user needs.

Enhancing Usability with User-Friendly Features
Industrial applications require intuitive interfaces that clearly communicate AI's reasoning. For instance, integrating explainability features in computer vision systems has improved operator trust and usability.

Learnings

Prioritizing the End-User
A user-centric approach is critical. Industrial partners emphasize designing XAI solutions based on the needs of operators rather than solely focusing on technological advancements. Engaging with end-users early and frequently ensures alignment with practical requirements. In addition, it is also important to facilitate continuous user engagement and feedback to ensure efficient human-AI collaboration.

There should be a balance between technical and human-centered focus. Focusing too much on AI's technical sophistication without considering user interaction can hinder adoption. A balanced approach, where AI models are designed with usability in mind, leads to better outcomes. With end users in mind, explainability should be tailored to the specific role of the user. Decision-makers may require high-level insights, while technical operators may need detailed justifications of AI recommendations.

Importance of Data Sources
The foundation of any successful industrial AI solution is domain-specific data. Process industries, which often act as customers, own these datasets. Beyond the technical aspects of organizing industrial data, ensuring access to relevant data while safeguarding customer rights is a critical challenge.
Key issues include:

- **Data Democratization:** The need to enable access to data within organizations to improve AI models while maintaining security and compliance.
- **Data Governance:** Defining roles and responsibilities among data stakeholders, including owners, engineers, and scientists.

As noted by Gröger [22] "There is no AI without data." The growing movement towards data-centric AI shifts the focus from model development to curating high-quality, well-structured data. This shift also underscores the importance of human-centric aspects, emphasizing that XAI solutions should be designed with end-user needs in mind.

[22] Gröger, C. (2021). There is no AI without data. *Communications of the ACM, 64*(11), 98-108. https://doi.org/10.1145/3448247

Extracting Useful Insights from Machine-Learning Models is Crucial
While AI models can make accurate predictions, their value lies in how actionable and interpretable their outputs are. Extracting useful insights from machine-learning models is essential for building trust and driving adoption. Consider explainability after solid AI performance is achieved.

Key Challenges

Despite the recognized benefits of HCXAI, integrating the principles generated from research into real-world AI application development remains a challenge.

Extracting and Understanding User Requirements
While recognizing the importance of human centric XAI, we also learn that gathering meaningful user requirements is often complex and time consuming, yet it is fundamental to developing relevant and practical XAI solutions that users would adopt.

- **Limited End-User Participation:** Recruiting experienced domain experts for requirement engineering and user testing is difficult compared to consumer product domains.
- **Complexity of Industrial Environments:** AI solutions must be tailored to highly specialized workflows, making user testing and validation more resource intensive. Testing prototypes in real working environments poses additional challenges, as operators for some industries can have high workloads during daily operations, which can result in limited chances for participation in trials and feedback sessions.

Complexity of AI Prediction Tasks
Some AI-driven tasks are inherently difficult, making explainability more challenging. Striking the right balance between model accuracy and interpretability is a persistent issue.

Integration with Legacy Systems and Existing Infrastructures
Many industries operate on legacy systems that are not designed for AI integration. The challenge lies in ensuring that XAI solutions can work seamlessly within existing infrastructures.

Managing Cost-Benefit Trade-offs
Investing in XAI must provide tangible benefits. Industrial partners stress the need for a well-defined strategy to ensure that the cost of implementing explainability features is justified by improved usability and decision-making.

Bridging the Gap Between Research and Deployment
Deploying AI solutions in industrial environments presents unique challenges, such as integration with legacy systems and operational constraints. Industry partners highlight the importance of transitioning from research prototypes to scalable production solutions.

Strategies to Overcome Challenges

Despite these challenges, there are some strategies to employ to mitigate their impact.

Using Structured User Research Methods
Employing established methods to gather user insights helps in developing XAI solutions that align with real-world needs.

Leveraging Ensemble Approaches
Combining multiple AI models or methodologies can enhance both accuracy and explainability, leading to more robust solutions.

Strengthening Collaboration with all key stakeholders
Collaboration is key to overcoming the challenges of developing explainable AI (XAI) solutions in process industries. Effective XAI development requires close cooperation among process industry stakeholders, AI developers, and domain experts. These partnerships will enable access to industrial data, ensure end-user involvement in the design process, and refine AI solutions to meet industry-specific needs.

Considering Traditional Alternatives
In some cases, non-AI solutions may still be viable. Traditional rule-based systems, statistical models, and human expertise can often provide sufficient decision support in certain scenarios. Additionally, leveraging well-established process optimization methods may be a cost-effective alternative to complex AI models. We stress the importance of critically assessing the specific problem at hand to determine if AI genuinely adds value, ensuring that resources are deployed in the most impactful way.

These learnings, gained from hands on design and development through this project, could guide organizations looking to navigate the complexities of AI adoption while ensuring transparency, trust, and usability.

9. Conclusion and Future Outlook

The development of Human-Centered Explainable AI (HCXAI) within process industries marks a transformative step toward making AI-driven decision-making more transparent, trustworthy, and effective for heavy industries. By emphasizing user needs, cognitive ergonomics, and domain-specific challenges, this guidebook has provided key insights and practical guidelines to designing explainable AI systems that align with human factors.

When this project began three years ago, Generative AI and Large Language Models (LLMs) were not as advanced as they are today. As a result, their potential roles in HCXAI have not been extensively explored in this guide. However, initial experiments suggest that these technologies could serve as intuitive and user-friendly interfaces to help users better understand machine learning models and outcomes. At the same time, challenges such as hallucination and user overreliance on Generative AI highlight the need for caution. Further research and experimentation are essential to fully understand how these tools can be responsibly integrated into explainable AI systems.

As AI continues to evolve, future developments in HCXAI will likely focus on adaptive explanations that cater to different user roles, real-time interactive explainability, and enhanced collaboration between humans and AI.

Emerging trends, such as multimodal explanations and personalized AI reasoning models, hold promise for further improving user trust and engagement. Additionally, regulatory frameworks and ethical considerations will continue to shape the landscape of explainable AI, requiring continuous refinement of best practices.

Abbreviations	Explanation
AI	Artificial Intelligence
XAI	Explainable AI
HCXAI	Human Centered XAI
ML	Machine Learning
UCD	User Centered Design
LLM	Large Language Model
MLOps	Machine Learning Operations

10. References

[1] European Parliament. (2023, August 6). *EU AI Act: first regulation on artificial intelligence.* https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#ai-regulation-in-europe-the-first-comprehensive-framework-4

[2] Guidotti, R., Monreale, A., Ruggieri, S., Pedreschi, D., Turini, F. & Giannotti, F. (2018). Local rule-based explanations of black box decision systems. *arXiv preprint* arXiv:1805.10820. https://doi.org/10.48550/arXiv.1805.10820

[3] Liao, Q. V. & Varshney, K. R. (2021). Human-centered explainable ai (xai): From algorithms to user experiences. *arXiv preprint* arXiv:2110.10790. https://doi.org/10.48550/arXiv.2110.10790

[4] Giovine, C., & Roberts, R. (2024, November 26). Building AI trust: The key role of explainability. *McKinsey & Company.* https://www.mckinsey.com/capabilities/quantumblack/our-insights/building-ai-trust-the-key-role-of-explainability?stcr=69646FBAFE6F499586EA699BC85CCCFD&cid=other-eml-alt-mip-mck&hlkid=88deba8a31c3417bb98c5c8026f9a57a&hctky=15573398&hdpid=e5d39fe2-0ab6-4f2f-a6fe-ab9561f99394

[5] Market and Markets. (2025, May 7). *Explainable AI Market worth $16.2 billion by 2028.* https://www.marketsandmarkets.com/PressReleases/explainable-ai.asp

[6] Di Bonito, L. P., Campanile, L., Di Natale, F., Mastroianni, M. & Iacono, M. (2024). eXplainable Artificial Intelligence in Process Engineering: Promises, Facts, and Current Limitations. *Applied System Innovation (ASI)*, 7(6). doi: 10.3390/asi7060121

[7] Hackos, J. T., & Redish, J. C. (1998). User and task analysis for interface design. *John Wiley & Sons, Inc.*.

[8] Naikar, N. (2016). *Work domain analysis: Concepts, guidelines, and cases.* CRC press.

[9] Zohrevandi, E., Brorsson, E., Darnell, A., Bång, M., Lundberg, J. & Ynnerman, A. (2023). Design of an Ecological Visual Analytics Interface for Operators of Time-Constant Processes. *2023 IEEE Visualization and Visual Analytics (VIS)*, Melbourne, Australia, 2023, pp. 131-135, doi: 10.1109/VIS54172.2023.00035.

[10] Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: informing design practices for explainable AI user experiences. *In Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-15). https://doi.org/10.1145/3313831.3376590

[11] Nauta, M., Trienes, J., Pathak, S., Nguyen, E., Peters, M., Schmitt, Y., ... & Seifert, C. (2023). From anecdotal evidence to quantitative evaluation methods: A systematic review on evaluating explainable ai. *ACM Computing Surveys*, *55*(13s), 1-42. https://doi.org/10.1145/3583558

[12] Silva, A., Schrum, M., Hedlund-Botti, E., Gopalan, N. & Gombolay, M. (2023). Explainable artificial intelligence: Evaluating the objective and subjective impacts of xai on human-agent interaction. *International Journal of Human–Computer Interaction*, *39*(7), 1390-1404. https://doi.org/10.1080/10447318.2022.2101698

[13] Hoffman, R. R., Mueller, S. T., Klein, G. & Litman, J. (2018). Metrics for explainable AI: Challenges and prospects. *arXiv preprint arXiv:1812.04608.* https://doi.org/10.48550/arXiv.1812.04608

[14] Dix, M., Koltermann, J., Mieck, S., Pastler, B. & Kloepper, B. (2024). XAI for anomaly analysis by power plant operators-a case and user study. *In ML4CPS–Machine Learning for Cyber-Physical Systems.* UB HSU.

[15] Lai, V., Chen, C., Smith-Renner, A., Liao, Q. V. & Tan, C. (2023). Towards a science of human-AI decision making: An overview of design space in empirical human-subject studies. *In Proceedings of the 2023 ACM conference on fairness, accountability, and transparency* (pp. 1369-1385). https://doi.org/10.1145/3593013.3594087

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

[17] Selvaraju, R. R., Lee, S., Shen, Y., Jin, H., Ghosh, S., Heck, L., ... & Parikh, D. (2019). Taking a hint: Leveraging explanations to make vision and language models more grounded. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 2591-2600).

[18] Kraus, M., Steinmann, D., Wüst, A., Kokozinski, A. & Kersting, K. (2024). Right on time: Revising time series models by constraining their explanations. arXiv preprint arXiv:2402.12921.

[19] Faubel, L., Woudsma, T., Kloepper, B., Eichelberger, H., Buelow, F., Schmid, K., ... & Bång, M. (2024). MLOps for cyber-physical production systems: challenges and solutions. *IEEE software*. https://doi: 10.1109/MS.2024.3441101

[20] Kreuzberger, D., Kühl, N. & Hirschl, S. (2023). Machine learning operations (mlops): Overview, definition, and architecture. *IEEE access*, *11*, 31866-31879. https://doi: 10.1109/ACCESS.2023.3262138.

[21] Public system developed by EXPLAIN project to build MLOps and XAI in Gitlab. (2025 May 7). https://www.uni-hildesheim.de/gitlab/explain/mlops-environment-for-explanatory-ml-models

[22] Gröger, C. (2021). There is no AI without data. *Communications of the ACM, 64*(11), 98-108. https://doi.org/10.1145/3448247

Main Contributors

Swedish consortium

ABB AB Corporate Research
Yanqing Zhang
Emmanuel Brorsson
Hugo Wärnberg
Dawid Ziobro
Gayathri Gopalakrishnan
Joakim Åström

Linköping University
Carl Westin
Elmira Zohrevandi
Magnus Bång
Kostiantyn Kucher

Södra Skogsägarna Ekonomisk Förening
Andreas Darnell

Viking Analytics AB
Arash Toyser
Rickard Claeson

Umeå University
Andreas Theodorou
Leila Methnani
Virginia Dignum

Boliden Mineral AB
Rasmus Tammia

German consortium

ABB AG Forschungszentrum
Ruben Huehnerbein
Pablo Rodriguez
Marcel Dix
Fabian Buelow
Nilavra Bhattacharya
Nika Strem
Sylvia Maczey
Matthias Ewald
Benjamin Klöpper (former employee)
Gianluca Manca (former employee)

Lausitz Energie Kraftwerke AG
Jan Jens Koltermann
Sebastian Mieck

Technische Universität Darmstadt
Maurice Kraus
Daniel Ochs

University of Hildesheim
Leornard Faubel-Teich
Klaus Schmid

Eraneos Analytics
Denis Baskan
Johannes Wagner

The Netherlands' consortium

Prodrive Technologies Innovation Services
Thomas Woudsma
Sjoerd van den Bos

Mek Europe
Henk Biemans
Shoich Rashed

Signify
Willem van Driel

Delft University of Technology
Amir Ghorbani

Acknowledgment

A heartfelt thank you to all members who contributed to the project, the users we interviewed, and the use case owners who supported us!

This project was supported by funding from VINNOVA Sweden (2021-04336), the German Federal Ministry of Research, Technology and Space (Bundesministerium für Forschung, Technologie und Raumfahrt, BMFTR; 01IS22030), and the Netherlands Enterprise Agency (Rijksdienst voor Ondernemend Nederland; AI212001). We gratefully acknowledge their support.

Other Project Members

Linköping University
Gustaf Söderholm
Annika Wilander
Adam Westergren
Danny Tran
Ebba Silfver
Erik Helmer
Joakim Andersson
Gustav Peterberg
Hans Birkedal
Jonas Lundberg

Viking Analytics AB
Rajet Krishnan
Adam Thörnblom
Axel Pantzare
Mohsen Nosratinia
Vishnu Nadhan
Sergio Martin Del Campo

Södra Skogsägarna Ekonomisk Förening
David Svahn
Andreas Eriksson
Thomas Håkansson
Linnéa Jakobsson

ABB AG Forschungszentrum
Martin W. Hoffmann
Matthias Biskoping

Lausitz Energie Kraftwerke AG
Erik Federau

Technische Universität Darmstadt
Kristian Kersting

University of Hildesheim
Jan-Henrik Böttcher
Carsten Wenzel

Eraneos Analytics
Daniel Meyer
Anne Ernst
Martin Schneider
Damian Gola
Steffen Maas

Signify
Rafal Jedrzejowski
Ronald Maandonks

Mek Europe
Erik van Reusel
Tomek Jasinski
Denis Angelov
Laine Mariquit
Jeremy Saise

Delft University of Technology
Justin Dauwels

Prodrive Technologies Innovation Services
Adam Dubowski
Pieter Derks

Date:
2025-05-12

Disclaimer
This document reflects only the views of its contributors. The funding agencies, project partners, and affiliated organizations are not responsible for any use that may be made of the information contained herein.

Usage Notice

- Logos and trademarks remain the property of their respective owners and are used with permission.
- Names of individual contributors are published with consent.
- No confidential, proprietary, or trade-secret information is disclosed in this document.
- This material may be shared for educational, research, and dissemination purposes in accordance with project agreements and funding guidelines.

Partners

ABB
mek marantz electronics ltd
BOLIDEN
LEAG
PRODRIVE TECHNOLOGIES
signify
SÖDRA
eraneos
VIKING ANALYTICS
LINKÖPING UNIVERSITY
TECHNISCHE UNIVERSITÄT DARMSTADT
TU Delft
Stiftung Universität Hildesheim 2003
UMEÅ UNIVERSITY

Funding support

Federal Ministry of Research, Technology and Space
VINNOVA Sweden's Innovation Agency
Rijksdienst voor Ondernemend Nederland
EXPLAIN

Date:
2025-09-12