Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 232 (2024) 2626–2638

Procedia Computer Science

www.elsevier.com/locate/procedia

5th International Conference on Industry 4.0 and Smart Manufacturing

Practical Aspects of Designing a Human-centred AI System in Manufacturing

Yuji Yamamoto[a]*, Alvaro Aranda Muñoz[b], Kristian Sandström[a]

[a]Mälardalen University, Bruksgatan 3B, Eskilstuna 63217, Sweden
[b]Research Institutes of Sweden, Stora Gatan 36, Västerås 72212, Sweden

Abstract

An increasing number of manufacturing companies have initiated designing and implementing AI systems in manufacturing, however, with limited success. Within our overarching research objective of establishing a methodology for the development of AI systems in manufacturing with socio-technical system consideration, this paper focuses on the early design phase of the development life cycle and aims to identify factors that are essential in the phase but whose importance has been less addressed in the manufacturing literature. To this aim, a case study was conducted adopting a design science approach. The case company was developing an ML-based anomaly detection system for a casting process. The researcher organised an AI system design workshop where participants from the company used the Human-AI design guidelines created by a leading large software company. The workshop enabled the participants to explore a wide range of design concerns. It, however, caused the confusing experience that they had to deal with too many questions simultaneously without clear guidance. Analysing this negative experience has led to identifying four design issues requiring further attention in the research. An example of these issues is that the interdependency of design decisions on operational procedures, human-machine interfaces, ML models, pre-processing, and input data makes it challenging to design these elements in isolation. The study found that a structured approach to dealing with the identified issues was currently lacking. This paper contributes to the manufacturing research community by addressing key unresolved issues in the research through highlighting practical details of designing AI systems in manufacturing.

Keywords: Human-centred AI; Manufacturing; AI system design; Machine Learning; Socio-technical systems

1. Introduction

Manufacturing companies recognise the vast potential of integrating AI-driven applications, hereafter called AI systems, in the manufacturing process. For instance, machine learning (ML) models can be used to monitor or predict

10.1016/j.procs.2024.02.081

manufacturing equipment conditions, detect anomalous patterns in manufacturing-related data and raise early signals to production personnel, or provide interpretable insights from high-dimensional data [1], [2]. An AI system can also help increase production planning agility by the ML model inferring more accurate future material demands, available resources, and lead times [3]. Further, a system can enable autonomous decision-making of multi-axis and transportation robots or augment or automate the human task of quality inspection on products or their components [2], [4].

On the other hand, the effort of adopting AI technology in manufacturing also uncovers various challenges. The use of the technology is often intended for high-stakes domains where faulty judgements, decisions, or controls cause serious negative impacts on productivity, quality, speed, and safety in manufacturing operations [2], [5]. High-stakes processes in manufacturing have been the subject of continuous improvement. Thus, the probabilistic behaviour of machine learning models may not easily surpass the achievement of the previous continuous improvement or the requirement or expectation from manufacturing experts.

Data availability and quality are often problematic in developing and deploying AI models in manufacturing [6]. Building a high-performance ML model often requires a substantial amount of structured data, which may not be available or difficult to obtain in practice. Oftentimes, a large part of manufacturing data relies on manual inputs, and the inputs are made without serious consideration of data quality critical for the model building [2], [6]. Further, in manufacturing, the data is often "dynamic"—the nature of data can asynchronously change over time by the changes in equipment, personnel, manufacturing processes and materials. ML model prototypes are often built on the data taken from a specific period of dynamic data [5].

The knowledge gap between manufacturing personnel and data scientists is also a significant challenge [6]. The former seldom have an in-depth understanding of ML techniques and their capabilities and limitations, which makes it difficult for the personnel to set proper expectations or specifications for AI systems. A few data scientists have deep insight into the reality of shop floor operations. This may lead to the development of an AI system being conducted with little proximity to the manufacturing organisation, failing to effectively utilise domain experts' knowledge, experience, and creativity [1]. Further, the current publicity of the potential and concerns of AI in society, such as the use of Large Language Models, causes fear or resistance to AI technology among those whose jobs may be affected by automation [7].

While the evidence of successful pilots or deployment of AI systems in manufacturing has been reported in the literature [6], the knowledge of methodological support to design and integrate AI systems into the manufacturing processes has been underdeveloped. While there are a few preceding studies [5], [8]–[11], they mostly focus on the needs and requirement identification phase of the system development life cycle. Outside the manufacturing domain, scholars and practitioners have proposed various design guidelines for AI systems e.g. [12], [13]. These artefacts provide concrete prescriptions for how to execute the design tasks. However, they are made mostly from data science, user experience, and software engineering perspectives and do not specifically address the manufacturing-specific context and challenges. Manufacturing operations are often understood as a socio-technical system. Considering how production personnel are involved in the AI system development and how they collaborate with the AI system in operations is particularly important in manufacturing [1], [11].

Our long-term research goal is to develop a methodology for designing and integrating AI systems in manufacturing with consideration of socio-technical system aspects. In this paper, we concentrate on an early design phase of the AI system development life cycle, which involves converting stakeholder desires, needs, and requirements into a system specification in broad terms. This paper aims to identify factors that need to be considered in this phase, hereafter called design issues, but are insufficiently addressed in the manufacturing literature.

To this purpose, we adopted design science [14] as the research approach. We conducted a case study where the company aimed to design and integrate an anomaly detection ML model into the manufacturing process. During the study, we organised an AI system design workshop based on a Human-centric AI (HAI) design guidebook developed by a leading large software company. The workshop successfully facilitated multi-disciplinary collaboration and enabled the workshop participants to explore a wide range of design concerns and their potential solutions. At the same time, however, the workshop caused the participants a confusing and disorganised experience. For instance, they felt overwhelmed by the multitude of design concerns they had to address simultaneously without clear guidance on prioritisation and resolution strategies. The analysis of this negative experience has led to the identification of four shortcomings of the HAI design guidelines when they are used in the manufacturing domain. Through further analysis of these shortcomings, the study has identified four design issues receiving insufficient attention in the current manufacturing literature. The main contribution of this paper is to provide practical insights into the design of an AI system in manufacturing. The identified issues can be considered for future research to construct an AI system development methodology for manufacturing.

2. Related research

2.1. AI systems in manufacturing

To further discuss the design of an AI system, it is essential to specify the construct of an AI system. In literature, an AI system in industry has been modelled in different ways. For instance, Lee [2] presents a conceptual framework of an industrial AI system that contains data, analytics, platform, and operation technologies as key enabling components. Rožanec et al. [15] present an AI system architecture consisting of multiple functional layers such as user interaction and data visualisation, data analytics, data processing, and data management. In this paper, we understand an AI system based on a model suggested by Kaymakci et al. [5]. Their simple model makes the discussion of the design of an AI system easier.

In the model, a data processing unit (DPU) is a base functional unit of an AI system and consists of a data input, transformation, and output module. An AI system includes three types of DPUs: pre-processing, ML model, and agent. A pre-processing DPU receives data from outside the AI system and processes them to be used by an ML model. An ML model is another type of DPU. An agent is the other type of DPU transforming the outputs from the ML model for further applications. Examples of an agent are a human-machine interface or a control unit to a piece of manufacturing equipment.

In this paper, we are interested in designing an AI system in manufacturing, considering socio-technical system aspects. We are particularly interested in two aspects. Firstly, we weigh the importance of people in manufacturing organisations actively participating in the process of designing and integrating AI systems in manufacturing. Their participation and unleashing their creativity and domain expertise are vital to utilising the full potential of technology adoption and ensuring sustainable changes toward high-performance operations [16]. The other aspect is to establish human-AI symbiosis. An AI system does not necessarily replace human manufacturing tasks but can augment them [17]. Bousdekis et al. [18] and Emmanouilidis et al. [11] present high-level conceptual models depicting how humans and AI can collaborate during manufacturing operations. In the models, an AI system assists operators in acquiring better situational awareness and more informed decision-making. At the same time, humans provide feedback to the system to improve the system's performance. Bousdekis et al. [19] further report a pilot case of human-AI symbiosis where an AI system provides operators dynamic work instructions for visual quality inspection based on the data from previous manufacturing processes, and the operators provide feedback to the system for better system performance. While the desired state of human-AI symbiosis in manufacturing environments has been envisioned in previous studies, it is scarecely discussed how such a system can be designed in practice. Moreover, how to methodologically enable production personnel's active participation in the process of AI system design and integration has been little discussed in the manufacturing literature.

2.2. AI system design for manufacturing

The design of an AI system is a pivotal phase in the AI system development life cycle. In the software engineering literature, the design phase denotes transforming the requirements for the system into a system description suitable for the implementation—coding with programming languages [20], [21]. At the highest level of abstraction, a system description is written in broad terms using the language of the problem environment. At lower levels of abstraction, a more procedural orientation is taken, for instance, describing the system in UML [21]. In this paper, our interest lies in an earlier design phase, meaning converting requirements into a system description in broad terms. We understand that the requirement specification and the early system design significantly interleave, and multiple iterations occur between these two phases [20].

In the manufacturing literature, there are a few previous studies aiming to structure the design of an AI system [5], [9]–[11], [22]–[24]. They are, however, scarce and in their infancy. In the first reference [5], the authors present a high-level AI system development life cycle comprising four stages: planning, experimentation, implementation, and operation. The early design phase is related to the planning stage. In each stage, this life cycle model suggests activities to be performed. For instance, at the end of the planning stage, one should develop a first draft of an AI system considering data, output functions, system performance, and usage environment. While such a suggestion provides some guidance, the granularity of the suggestions in the model is generally low for operationalisation. Further, the model exhibits little consideration for the participation of production personnel in the development process and establishing human-AI symbiosis.

The other references ([9]–[11], [22], [24]) are from the same research project called the STAR project. The objective of this project is consistent with that of our long-term research goal—developing a methodology for designing and implementing an AI system in manufacturing. The STAR project also emphasises stakeholder participation in the development and establishment of human-AI symbiosis. It identifies three phases of AI system development: definition and design, early development and testing, and final development and testing. Currently, the methodology for the first phase has been reported. In this phase, stakeholders and AI technology experts collaboratively identify user and system requirements from the perspectives of the human-AI symbiosis and the desired system's impact on operational performance. Other perspectives related to the symbiosis, such as the system's trustworthiness, safety, and explainability, and user feedback to the system, are considered in the requirement elicitation [24]. Then, "technical components" pertinent to these requirements, for instance, collaborative robot, camera, PLC, and human-machine interface, are identified. An initial design of the AI system is accomplished by linking those requirements and components. A flow chart illustrating the potential collaboration between humans and the AI system is also depicted.

The reported results from this project are an important precedent for our work. However, while their requirement elicitation process deals with a wide range of human-AI-related issues (see Table 1), the effectiveness of the early design is unclear. For instance, it is unclear how the system's explainability can be addressed only by identifying technical components or drawing a high-level flow chart.

2.3. AI system design for unspecified domains

Not limited to the manufacturing domain, there is an increasing body of literature presenting and discussing human-centric AI (HAI) design guidelines [13], [25]–[28]. These guidelines are a consolidation of the theories and practices in design thinking, user experience (UX), and responsible AI [12]. Several large software companies have published HAI guidelines. For instance, Microsoft has proposed 18 design guidelines for human-AI interaction [29]. Examples of the guidelines are: help the user understand what the AI system is capable of doing; enable the user to provide feedback indicating their preferences during regular interaction with the AI system. The People+AI (PAIR) research team at Google published a design guidebook that facilitates designers and users to co-design an AI system [13]. The PAIR guidebook includes 23 design guidelines to address key design concerns. Further, Subramonyam et al. [12] analysed existing HAI design guidelines, in total 249 guidelines, and found that guidelines could be classified into one of four categories: 1) identification of user expectation and needs, 2) design of user interface, 3) design of AI models, and 4) considerations for collection, processing, and evaluation of data.

These design guidelines are relevant to our study because they focus on the design phase of the AI system development and intend to facilitate stakeholder involvement and the effectuation of human-AI symbiosis. The guidelines cover a wide range of human-AI design concerns and provide concrete prescriptions for how to address them. Further, they often offer examples of how those concerns have been addressed in real-world applications. However, the validity of these practitioner-oriented guidelines for the manufacturing domain is uncertain. For instance, in the study presented by [29], the usefulness of the proposed 18 design guidelines was evaluated by the developer of AI-powered software products, such as email clients, social-network applications, and music players [29]. These applications are, however, less relevant to manufacturing applications.

To summarise, the review of the related previous work indicates various factors that need to be considered in the early design of AI systems. We believe that the methodology reported from the STAR project [9]–[11], [22], [24] and the HAI guidelines [13], [25]–[28] are state-of-the-art. Factors addressed in those works are summarised in Table 1.

Generally, the method reported from the STAR project [9]–[11], [22], [24] addresses more manufacturing-specific concerns, but it is mostly related to requirement specification and lacks practical design guidelines. The HAI guidelines are equipped with richer practical design guidance but are less exposed to the manufacturing domain.

Table 1. Design issues covered by AI system design methods.

	The AI system design method for manufacturing — a result of the STAR project [9]–[11], [22], [24].		The PAIR guidelines [13] as a representative set of HAI guidelines.	
Design issues:	Has explicit guides to discuss the issue in requirement elicitation	Has explicit guides to address the issue in the early design	Has explicit guides to discuss the issue in requirement elicitation	Has explicit guides to address the issue in the early design
Human-AI collaboration workflow		x		
Balance of control and automation.	x		x	x
Explainability of the system	x		x	x
Inference error handling	x		x	x
Feedback to the AI system for AI to learn	x		x	x
How to present inference results to users	x		x	x
Onboarding of users for the system introduction			x	x
Data needs, collection, and labelling			x	x
Security, safety, and privacy	x		x (for data privacy and ethics)	x (for data privacy and ethics)
AI model and system performance	x (for the system performance)		x	x

3. The research method

The review of the previous studies identifies several important factors that need to be considered in the early phase of the AI system design, as shown in Table 1. It is unlikely that these design issues encompass all the considerations, considering the immaturity of the research in this field. To further explore and identify factors that may have been less regarded in the previous work, we conducted an empirical study adopting the design science approach [14].

Design science is a specific type of action research. In action research, researchers are active participants in the process of change in the case study [30]. Researchers in design science are particularly interested in creating knowledge by developing and implementing solutions to intervene in the phenomena under study [14]. Design science is a suitable approach to tackling an ill-structured problem—a solution to a problem is not known to decision-makers or does not even exist [31]. We adopt the design science approach for two reasons. First, how to design an AI system in manufacturing is poorly understood in the industry and thus is an ill-structured problem. Second, adopting the approach enabled an industry-academia collaboration. There, we researchers would identify or develop an AI system design method and experiment with it at the case company, and the company would develop an AI system with the help of the method. Through this collaboration, the researchers could pursue the research inquiry stated in this paper.

Design science inherits typical risks of action research, such as difficulty in achieving reproducibility and researchers' bias in interventions. Moreover, researchers in design science need to take risks of failing to create effective solutions for the case company [14]. To undermine these risks, the study was conducted by three researchers specialising in new technology adoption in manufacturing, industrial AI system development, and user experience. These researchers had several years of practice experience in the respective areas. This multi-disciplinary research team and close collaboration with practitioners at the case company could create intersubjectivity, which would mitigate the risk of researcher bias.

The case company is a large company manufacturing power distribution equipment. The case study was conducted at one of the manufacturing sites in Sweden. The site was developing an ML model that would detect anomalous data patterns from temperature sensors installed in the multiple furnaces in the casting process. The company developed a prototype of the ML model and wanted to design an AI system, especially the agent part of the system.

The research team considered this company's context as an opportunity to understand what design issues the current state of the art could address and not address. The team decided to use one of the HAI design guidelines, the People+AI (PAIR) Guidebook [13]. The design method proposed in the STAR project [9]–[11], [22], [24] was excluded because it mainly focused on requirement identification and the agent part of the AI system design was less addressed than the HAI guidelines.

There were other alternatives for the HAI guidelines than the PAIR guidebook, which were Microsoft's "HAX toolkit" [25], [29], Apple's "Human Interface Guidelines" [27], IBM's "AI Essentials Framework" [26], and "AI meets design toolkit" [28]. The PAIR guidebook was chosen because, comparatively, it contained more design support, including practical examples from the industry and guidance on how to organise design workshops focusing on facilitating stakeholders' active involvement in the design process [32].

Based on the workshop guidance provided by the PAIR guidebook, a one-day design workshop was organised at the case company in February 2023. The research team facilitated the workshop, and ten practitioners from the case company attended the workshop, whose roles were R&D engineers, process engineers, data scientists, manufacturing technicians, and Lean Six Sigma experts. The goal of the workshop was for the participants to firstly understand the opportunities and challenges that they might face when they would integrate the anomaly detection ML prototype into the shop floor, and secondary, create a basic design of the agent part of the AI system addressing the identified opportunities and challenges.

The workshop was planned to include the following sessions. 1) a warm-up session to generally think about creating human-centred AI systems by examining an industrial use case, 2) focusing on the AI system to be developed, discussion on how the system should present the result of the ML inferences to users and how it should interact with them, 3) exploration on problematic situations that could occur during the user interactions, and identification of the possible solution plans to those problems, 4) detailing the solutions plans from pre-defined perspectives, for instance how the system should recover from the errors, and how to establish the balance between automation and user control of the system.

While the researchers facilitated the workshop, they observed and listened to the participant's behaviours and conversations. The conversations were audio recorded. At the end of the workshop, a reflection session was held among the researchers and participants. The researchers were aware of the generality of the single case study. The generality was sought by the researcher seeking transferability—the description of the context and finding from the case is detailed enough for the readers to assess the generality of the cases or make comparisons with their own or other reported situations [30].

4. Result

The workshop resulted in conducting the first three sessions instead of the planned four described in Section 3. It was found that the participants became more confused and cognitively overloaded as the session progressed, and they thought less meaning to move on to the fourth session. This section describes what was discussed and observed in these three sessions.

The first session: A warm-up for thinking about creating human-centred AI

The purpose of this session is for the participants to understand key considerations in the design of a human-centred AI system. The participants watched a use case video supplied by the PAIR guidebook and reflected on human-AI symbiosis aspects of the AI system they wanted to develop.

Even though it was a warm-up session, the participants' reflections captured a wide range of symbiosis aspects. For instance, the participants discussed the risk of the AI system being too much trusted by operators. Those who have rich experience in the process could verify AI results by looking at the underlying data, but this might not be possible for those less experienced. A participant said that the AI system should present additional information that could explain the model behaviour, for instance, a confidence score of the inference and the time-series data on which the anomaly was detected. Another participant discussed that the probabilistic behaviour of the AI system should be carefully explained to the users when it is introduced to them since the behaviour would be different from those of traditional manufacturing equipment and IT systems. Another participant said that operational producers triggered by the anomaly alarm would be important to deal with the uncertainty of the system's behaviour. Participants also discussed that the economic impact would be much larger in their use case when the AI system did not work properly than the one shown in the video—an AI-driven application designed to assist biomedical researchers in conducting effective literature searches.

The second session: discussion of how the AI results should be presented to users

In this session, the participants discussed how the ML results should be presented to users and how the system should interact with them. This session was a necessary step before the third session where the participants would discuss potential problems in the interactions and a rudimentary plan to resolve them.

It was discussed that the workers operating the casting process would first see the alarms. The workers then would look at the information screen in the system and decide on further actions. Actions could be the workers evaluating the situation by themselves or contacting process engineers for deeper evaluation and decision-making. Equipment suppliers or service companies could be contacted depending on the result of the diagnosis. The information presented in the system, therefore, should be able to be shared among those internal and external actors.

A group of participants worked on a paper prototype that showed what information should be presented in the system's interface. Different ideas were generated. For instance, one idea was that the system should show an overview of the shop floor of the casting process and on which furnaces the system detected anomalous signals. A participant said that operators should be able to see, for instance, the machine's ID number on which an anomaly was detected, which signal was flagged as anomalous, a description of the mode of anomaly, and historical time-series data annotated with what actions were taken on the previous alarms. To gain this historical data, operators and process engineers should be able to provide feedback to the system about what actions were taken in response to the alarms. This information could also help data scientists to improve the ML model after the deployment.

While various ideas of the interface design were discussed in this session, the participants found it difficult to specify or detail the design further because it would depend on various factors, such as the mode of anomaly, the degree of its urgency, the accuracy of the inferences, knowledge and experience of operators, and role and responsibility of those who would interact with the system. Due to these multi-faceted factors, it was found that an anomaly alarm could lead to multiple patterns of operational procedures conducted by different actors. It was cognitively stressful for the participants to imagine all those potential patterns of procedures when considering the system interface.

The third session: discussions of problems and solutions at critical moments using the design guidelines

The PAIR guidebook defines critical moments as those when humans interact with an AI-driven system and when their trust in the system can be earned or lost [13]. The researchers selected the following critical moments for the discussion: when an alarm is raised to the operators, when operators or actors provide feedback to the system, when maintenance of the algorithms becomes necessary due to the changes made in the casting process, and when the AI system is introduced to operators. The purpose of the third session is, by having these critical moments as the reference points, the participants discuss problems that could occur in those moments and identify solutions to those problems.

To facilitate the discussion, the participants consulted the design guidelines provided by the PAIR guidebook. Examples of guidelines are; "Be transparent with your users about what your AI-powered product can and cannot do" and "If you decide to show model confidence, make sure it's done in a way that's helpful to your users".

The participants discussed various possible problems. For instance, too many false positives might cause alarm fatigue among operators; frequent alarms might lead to operators or engineers spending too much time dealing with those alarms; operators might mistrust the system when the system's capability, behaviours, and limitations are not appropriately explained to them during the system introduction; presenting too much information might overwhelm the operators; how to detect false negative was uncertain; the algorism maintenance would be necessary when the nature of the data would be changed by, for instance, replacing sensors.

To address those potential problems, different measures were discussed. For instance, against the problem of too much information being presented to the operators, it was discussed that different levels of information should be presented for different roles, such as operators, process engineers, and problem-solving task force teams. For the problem of ML model maintenance, a participant suggested holding regular meetings with data scientists and production personnel to discuss the model maintenance issues.

It was observed that the session encouraged the participants to broadly explore potential problems and solutions regarding the interaction between humans and the AI system. The participants, however, felt that this session was even more mentally exhausting than the previous session because most of the discussions were of their imagination. They discussed the hypothetical situations related to the critical moments and further imagined possible problems and solutions without knowing the system's exact specifications and resulting behaviours. They felt that their design questions kept expanding without converging.

The participant's reflection on the workshop:

The participants experienced that the workshop enabled them to broadly explore multifaceted aspects of integrating the ML model into the manufacturing process. On the other hand, they found the workshop difficult, overwhelming, and confusing. They experienced that they had to deal with too many questions or design concerns at the same time, without much knowing how to answer or address them and in which order. Throughout the workshop, the number of unanswered questions and unresolved design concerns kept growing. The following comments from the participants represent their experience. A participant commented, *"It's difficult to implement human aspects because it means changes to humans. Even though we did the pre-study and developed an ML prototype, and the result looks great, I now see a huge mountain to climb. I expected that before, but now, standing by the mountain, I realise how huge it is."*. Another continued, *"If you compare this to building a house, in the pre-study, we put up wallpaper, and it looked interesting, but we haven't seen the design of the house, and have not decided on other parts, such as the foundation and walls. We need a better plan that enables us to tackle all these different parts in the right order."*

5. Analysis

The case study showed that the PAIR guidebook enabled the multi-disciplinary team to collaborate and explore requirements and potential solutions of the AI system from the perspective of human-AI symbiosis. At the same time, the participants experienced that designing solutions was particularly challenging, as they were faced with numerous design concerns and questions simultaneously, lacking clear guidance on how to address them in which priority. This section firstly analyses the reasons behind this experience, with the purpose of identifying the shortcomings of the design guidelines. The analysis identifies four shortcomings. While the study utilised the PAIR guidebook as a representative set of HAI guidelines, we consider that this analysis result is applicable to other HAI guidelines mentioned in Section 3 because of the similarity among those sets of guidelines. After this analysis, we also analyse how these shortcomings are addressed in the design method from the STAR project [9]–[11], [22], [24], which we considered another state-of-the-art design method for manufacturing.

5.1. Shortcomings of the HAI design guidelines

Our analysis has identified the following four shortcomings of the HAI design guidelines.

The first shortcoming: Designing the details of the operational procedures that integrate the AI system in different use scenarios was a vital part of the solution design, but methodological assistance was lacking.

We analyse that these operational procedures are key design elements when developing an AI system in manufacturing, especially when humans are involved in the system operations. Fig.1 shows five groups of design elements: input data, pre-processes, ML models, agents, and operational procedures. This figure is derived from the previous work [5], where the authors identified the first four groups. In the figure, we incorporate operational procedures as an additional group of design elements.

In the workshop, the participants suffered from the lack of structured means to make those procedures explicit. In the second session of the workshop, they had difficulties in detailing and deciding on the interface design because it was unclear about possible operational procedures in different usage scenarios. An anomaly alarm from the AI system could cause multiple patterns of operational procedures that could be affected by multiple factors, such as the mode and urgency of the alarm and the knowledge and experience of the operators and other actors. Relevant information for those involving actors could be different at different interaction points in different branches of procedures.

The mentioned alarm was one of the potential triggers activating different procedural pathways. The participants discussed other situations triggering more procedures, such as operators identifying false negatives or replacing degraded sensors with new ones, rendering previous data less relevant. In the workshop, participants used their imaginative abilities to depict diverse scenarios and the corresponding work procedures, which they found cognitively stressful. We analyse that this partially contributed to the participants' overwhelming and confusing experience.

The second shortcoming: Fig 1. shows five groups of design elements. It was found in the case study that decisions on design elements in one group often significantly impacted those in another group. This inter-influence, also depicted in Fig.1, resulted in substantial challenges and increased complexity for the participants making design decisions. However, the design workshop lacked a structured approach to deal with this interplay of influence.

The inter-influence can be further explained with examples. In the case study, we observed that the design of the human-machine interface (an instance of an agent) was significantly affected by the design of operational procedures. Influence in the opposite direction may happen in other use cases— the design of the interface affects the design of operational procedures.

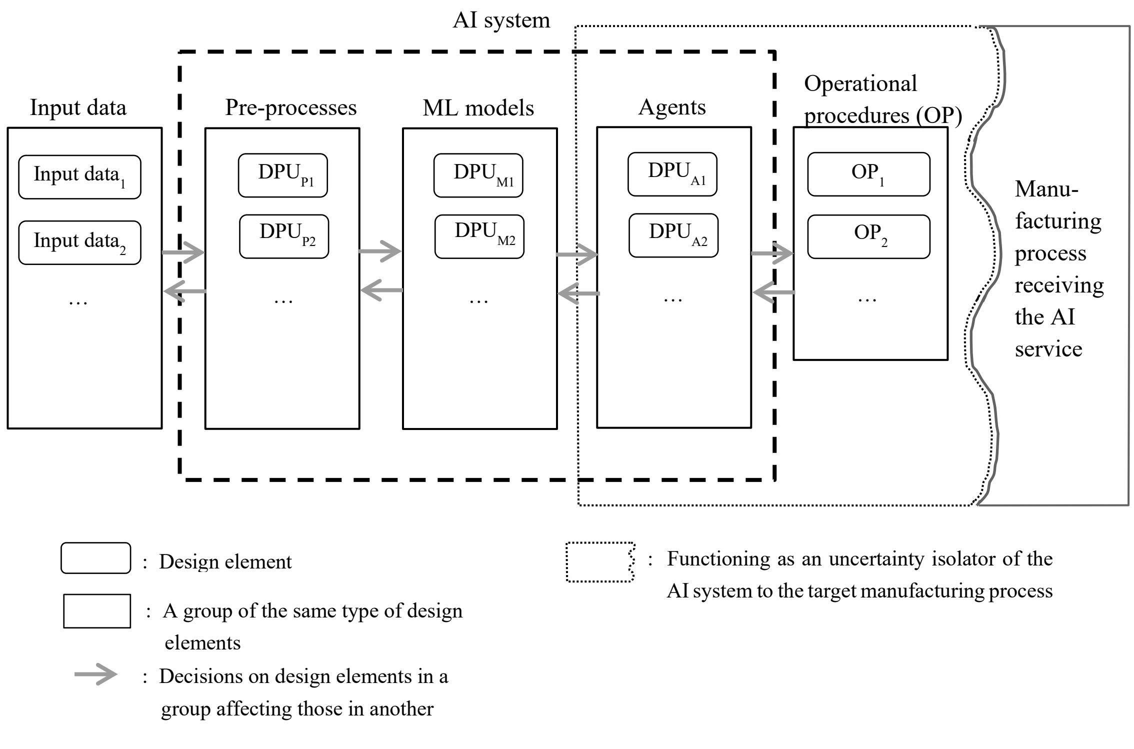

Fig. 1. Illustration of how decisions on design elements in a group affect those in other groups

The case study indicated that the interface design would be affected by the ML model behaviours and performance, which would be a result of the ML model design. The latter would be further affected by decisions on input data preprocessing and input data collection. We assume that the design influence in the opposite direction could happen in other use cases. For instance, the selection of an ML model could dictate the design of pre-processing and the choice of input data.

The workshop participants in the case study suffered from this inter-influence. For instance, in Session 3, the participants discussed one of the questions that the PAIR guidebook suggested addressing—how to build and calibrate users' trust in the AI system [13]. The participants quickly realised that the question would be largely affected by how the user interface was designed, which was found to be dependent on the design of the operational procedures in multiple scenarios. It was also found that the interface design was also affected by the ML model design and its performance, which would depend on the data and how it would be pre-processed. The participants experienced that a design-related question imposed a cascade of additional questions that could hardly be answered simultaneously. At the same time, they were not equipped with a structured way to handle these cascading questions. We analyse that this lack of means to handle the cascading questions was another source of the participants' confusing experience. They felt that design concerns were snowballing rather than converging into resolutions.

The third shortcoming: The importance of designing operational procedures raises the question of who is responsible for this task and what competence is required for it. The HAI guidelines did not address this question, as their focus was not on operational procedure design. Moreover, it became apparent in the case study that the HAI guidelines were primarily structured from the perspective of UX designers, who appeared to be responsible for consolidating the requirements and solutions identified from the workshop. The unclarity regarding who should be accountable for the design of operational procedures led to additional confusion among the workshop participants. It was uncertain if the accountable person would be the UX designer, operators, their leaders, technicians, or the project leader for the system development.

For other design elements than the operational procedures, the responsibilities are more clearly recognised in the literature. UX designers are typically responsible for the agent design when a user interface is involved, data scientists for ML models and pre-processing, and data engineers for incoming data and pre-processing [2], [5]. Kutz et al. [33] imply that domain experts should be responsible for designing operational procedures, but the term “domain experts” is still broad and lacks specificity. Nonetheless, the case study indicated that people responsible for the design of operational procedures should possess a deep understanding of shop floor operations at the same time, sufficient knowledge about other design elements. We recognise that the latter knowledge is particularly crucial because of the inter-influence discussed in this section.

The fourth shortcoming: As illustrated in Fig.1, it became evident through the case study that the agent and operational procedure design would be crucial in absorbing the uncertainties inherent in an AI system and connecting the AI system to often high-stakes, high-precision and low-variance manufacturing processes. The case study evidenced that the HAI design guidelines failed to address this crucial aspect.

Uncertainties in ML model outputs due to their probabilistic nature have been well recognised [29]. In a manufacturing context, the uncertainty can be magnified due to the dynamic nature of input data. The agents and operational procedures, which are usually designed based on rule-based logic, should act as an uncertainty isolator, bridging the two realms of uncertainty and certainty. We analyse that the lack of attention to the careful design of uncertainty isolator in the HAI guidelines is because those design guidelines are primarily tailored to, and draw their design knowledge from, applications in relatively lower-stakes domains, such as social media applications and music players.

5.2. Does the design method tailored to manufacturing contexts address these shortcomings?

This paper discusses that the results of the STAR project are another state-of-the-art in terms of the early design of AI systems for manufacturing. In this subsection, we analyse whether the identified shortcomings of HAI guidelines can be resolved by the design method suggested by the SATR project.

The first shortcoming discussed in this section was the absence of a systematic approach to designing operational procedures in different usage scenarios. The shortcoming can be mitigated to some extent through the method suggested by the STAR project. In the method, design participants depict a process flow that illustrates how the AI system can collaborate with humans during operations. This process flow, however, represents only an initial idea of the basic functionality of the AI system. The current state of the method does not encompass the detailed design of operational procedures required for various usage situations, which became necessary in the case study.

The second shortcoming—the lack of a structured approach to address the cascading design questions—remains unresolved with the method. As mentioned in Section 2, the method does not assist further design activities beyond connecting requirements with physical or functional components such as PLC and human-machine interface. Cascading questions often arise when delving into the details of the system design.

Regarding the third limitation, which involves the unclear allocation of responsibility for designing operational procedures, the method hardly addresses this issue. The method suggests that a multidisciplinary team should create the mentioned process flow. However, it remains uncertain whether the team continues to detail operational procedures in different usage scenarios and still holds determining responsibility for it.

Lastly, the method scarecely addresses the fourth shortcoming—the lack of emphasis and a structured approach to designing uncertainty isolators for high-stakes processes. In the method, the importance of designing such isolators may be captured in the requirement identification phase. It, however, does not equip with practical recommendations on how to design those isolators. The lack of structured knowledge about designing isolators may be an account for many AI initiatives in manufacturing adopting the "AI-first design process"[12] and later facing significant difficulties in integrating the AI system into high-stakes processes.

6. Conclusions and discussions

The goal of our research is to develop a methodology for designing and integrating AI systems into manufacturing processes, with consideration of socio-technical system aspects. The study presented in this paper focuses on the early design phase of the AI system development life cycle and aims to identify important factors that need to be considered in the design phase but are still insufficiently addressed in the manufacturing literature. The study revealed the four shortcomings of the HAI guidelines when they were applied in the manufacturing context. It is further analysed that the AI system design method tailored to the manufacturing context [9]–[11], [22], [24], which we considered state-of-the-art in the manufacturing literature, also fails to address the design issues causing those shortcomings. Therefore, we argue that these design issues are the knowledge gap requiring further attention in future research. Effectively addressing those issues should be critical when researchers or practitioners construct a comprehensive AI system development methodology.

There are at least two limitations in this study. First, we do not claim that the identified issues uncovered all the less-regarded design challenges in the current manufacturing research. The found issues are based on a single case study involving one design event for a specific use case. The main focus of the case study was on the design of agents and operational procedures and not on the other design elements. Future studies should aim to identify additional issues that may exist. Second, due to the mentioned research design limitation, the validity of the findings should be examined further in future studies. The importance of identified issues can vary depending on use cases. For instance, in other use cases, the importance of designing operational procedures may be less critical compared to the case study presented in this paper. At least, this paper contributes to raising awareness about the found issues for the researchers and practitioners interested in structuring a robust design method for an AI system in manufacturing.

Resolving the identified issues is not the scope of the present study but a topic for future study. The literature already offers various hints and approaches that can assist in tackling these issues. For instance, the concurrent design of operational procedures and user interfaces can be assisted by drawing swimlanes [33], using role-playing, or business process origami [34]. "Leaky abstractions"—tangible objects facilitating multidisciplinary collaboration [12] can be actively sought and used to mitigate the challenge of cascading design questions. Adopting an agile and prototype-driven approach is also inevitable to deal with this challenge.

7. References

[1] E. E. Makarius, D. Mukherjee, J. D. Fox, and A. K. Fox, 'Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization', *J Bus Res*, vol. 120, 2020, doi: 10.1016/j.jbusres.2020.07.045.

[2] J. Lee, *Industrial AI: Applications with Sustainable Performance*. Springer, 2020.

[3] L. Wang, Z. Pan, and J. Wang, 'A Review of Reinforcement Learning Based Intelligent Optimization for Manufacturing Scheduling', *Complex System Modeling and Simulation*, vol. 1, no. 4, pp. 257–270, 2021, doi: 10.23919/CSMS.2021.0027.

[4] J. Chen, Z. Zhang, and F. Wu, 'A data-driven method for enhancing the image-based automatic inspection of IC wire bonding defects', *Int J Prod Res*, vol. 59, no. 16, 2021, doi: 10.1080/00207543.2020.1821928.

[5] C. Kaymakci, S. Wenninger, and A. Sauer, 'A Holistic Framework for AI Systems in Industrial Applications', in *Lecture Notes in Information Systems and Organisation*, 2021. doi: 10.1007/978-3-030-86797-3_6.

[6] J. F. Arinez, Q. Chang, R. X. Gao, C. Xu, and J. Zhang, 'Artificial Intelligence in Advanced Manufacturing: Current Status and Future Outlook', *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, vol. 142, no. 11, 2020, doi: 10.1115/1.4047855.

[7] A. Dabbous, K. Aoun Barakat, and M. Merhej Sayegh, 'Enabling organizational use of artificial intelligence: an employee perspective', *Journal of Asia Business Studies*, vol. 16, no. 2, 2022, doi: 10.1108/JABS-09-2020-0372.

[8] S. Vernim, H. Bauer, E. Rauch, M. T. Ziegler, and S. Umbrello, 'A value sensitive design approach for designing AI-based worker assistance systems in manufacturing', in *Procedia Computer Science*, 2022. doi: 10.1016/j.procs.2022.01.248.

[9] C. Emmanouilidis, S. Waschull, J. A. C. Bokhorst, and J. C. Wortmann, 'Human in the AI Loop in Production Environments', in *IFIP Advances in Information and Communication Technology*, 2021. doi: 10.1007/978-3-030-85910-7_35.

[10] S. Waschull and C. Emmanouilidis, 'Development and application of a human-centric co-creation design method for AI-enabled systems in manufacturing', in *IFAC-PapersOnLine*, 2022. doi: 10.1016/j.ifacol.2022.04.246.

[11] C. Emmanouilidis and S. Waschull, 'Human in the loop of AI systems in manufacturing', in *Trusted Artificial Intelligence in Manufacturing: A Review of the Emerging Wave of Ethical and Human Centric AI Technologies for Smart Production*, 2021. doi: 10.1561/9781680838770.ch9.

[12] H. Subramonyam, J. Im, C. Seifert, and E. Adar, 'Human-AI Guidelines in Practice: Leaky Abstractions as an Enabler in Collaborative Software Teams', 2022.

[13] Google, 'People + AI Guidebook'. Accessed: Jun. 24, 2023. [Online]. Available: https://pair.withgoogle.com/guidebook

[14] J. Holmström, M. Ketokivi, and A. P. Hameri, 'Bridging practice and theory: A design science approach', *Decision Sciences*, vol. 40, no. 1, 2009, doi: 10.1111/j.1540-5915.2008.00221.x.

[15] J. M. Rožanec *et al.*, 'Human-centric artificial intelligence architecture for industry 5.0 applications', *Int J Prod Res*, 2022, doi: 10.1080/00207543.2022.2138611.

[16] L. Hannola, A. Richter, S. Richter, and A. Stocker, 'Empowering production workers with digitally facilitated knowledge processes – a conceptual framework', *Int J Prod Res*, vol. 56, no. 14, pp. 4729–4743, 2018, doi: 10.1080/00207543.2018.1445877.

[17] D. Romero *et al.*, 'Towards an Operator 4.0 Typology: A Human-Centric Perspective on the Fourth Industrial Revolution Technologies', in *In International conference on computers & industrial engineering (CIE46)*, Tianjin, China, Oct. 2016, pp. 1–11.

[18] A. Bousdekis, D. Apostolou, and G. Mentzas, 'A human cyber physical system framework for operator 4.0–artificial intelligence symbiosis', *Manuf Lett*, vol. 25, pp. 10–15, 2020.

[19] A. Bousdekis *et al.*, 'Human-AI Collaboration in Quality Control with Augmented Manufacturing Analytics', in *IFIP International Conference on Advances in Production Management Systems*, Springer, 2021, pp. 303–310.

[20] I. Sommerville, *Software engineering* , Tenth edition. Boston: Pearson, 2016.

[21] D. Sundar, *Software engineering* , First edition. New Delhi, India: University Science Press, an Imprint of Laxmi Publications Pvt. Ltd., 2010.

[22] C. Ipektsidis and J. Soldatos, 'User Studies and Co-Design for Human-Centred AI', 2021.

[23] C. Ipektsidis and J. Soldatos, 'Reference Scenarios and Use Cases for AI in Manufacturing', 2021.

[24] C. Ipektsidis and J. Soldatos, 'Report on Co-Design Workshops and Focus Groups-Initial version', 2021.

[25] Microsoft, 'The HAX Toolkit'. Accessed: Jun. 25, 2023. [Online]. Available: https://www.microsoft.com/en-us/haxtoolkit/

[26] IBM, 'AI Essentials Framework'. Accessed: Jun. 25, 2023. [Online]. Available: https://www.ibm.com/design/thinking/page/courses/AI_Essentials

[27] Apple, 'Human Interface Guidelines'. Accessed: Jun. 25, 2023. [Online]. Available: https://developer.apple.com/design/human-interface-guidelines/machine-learning

[28] N. Piet, ‘AI MEETS DESIGN TOOLKIT’. Accessed: Jun. 25, 2023. [Online]. Available: https://aixdesign.co/toolkit

[29] S. Amershi *et al.*, ‘Guidelines for human-AI interaction’, in *Conference on Human Factors in Computing Systems - Proceedings*, 2019. doi: 10.1145/3290605.3300233.

[30] R. Westbrook, ‘Action research: a new paradigm for research in production and operations management’, *International Journal of Operations & Production Management*, vol. 15, no. 12, pp. 6–20, 1995.

[31] H. A. Simon, ‘Does Scientific Discovery Have a Logic?’, *Philos Sci*, vol. 40, no. 4, 1973, doi: 10.1086/288559.

[32] J. Kutz, J. Neuhüttler, K. Schaefer, J. Spilski, and T. Lachmann, ‘Generic Role Model for the Systematic Development of Internal AI-based Services in Manufacturing’, in *the 56th Hawaii International Conference on System Sciences*, 2023, pp. 909–917.

[33] M. P. Sendak *et al.*, ‘Real-world integration of a sepsis deep learning technology into routine clinical care: Implementation study’, *JMIR Med Inform*, vol. 8, no. 7, 2020, doi: 10.2196/15182.

[34] Á. Aranda Muñoz, Y. Eriksson, Y. Yamamoto, U. Florin, and K. Sandström, ‘To support IoT collaborative expressiveness on the shop floor’, in *Proceedings of the Design Society*, 2021. doi: 10.1017/pds.2021.576.