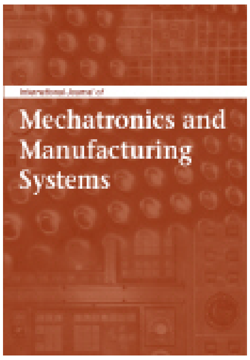

International Journal of Mechatronics and Manufacturing Systems

ISSN online: 1753-1047 - ISSN print: 1753-1039
https://www.inderscience.com/ijmms

Challenges in designing a human-centred AI system in manufacturing

Yuji Yamamoto, Aranda Muñoz Álvaro, Kristian Sandström

DOI: 10.1504/IJMMS.2024.10068969

Article History:

Received:	21 February 2024
Last revised:	17 June 2024
Accepted:	10 September 2024
Published online:	04 February 2025

Challenges in designing a human-centred AI system in manufacturing

Yuji Yamamoto*

School of Innovation, Design, and Engineering,
Mälardalen University,
Bruksgatan 3B 632 17 Eskilstuna, Sweden
Email: yuji.yamamoto@mdu.se
*Corresponding author

Aranda Muñoz Álvaro

Research Institutes of Sweden,
Stora Gatan 36, Västerås, 72212, Sweden
Email: alvaro.aranda.munoz@ri.se

Kristian Sandström

School of Innovation, Design, and Engineering,
Mälardalen University,
Bruksgatan 3B 632 17 Eskilstuna, Sweden
Email: kristian.sandstrom@mdu.se

Abstract: Despite successful AI system deployments in manufacturing, methodological support for developing and integrating AI systems into manufacturing processes remains underdeveloped. This paper aims to identify gaps in the methodological support for the early design phase of AI system development in manufacturing. The study reveals the thinking-level challenges that design participants face in the early design phase and identifies remedies for those challenges, which are only superficially addressed in current manufacturing literature. The paper contributes to uncovering the current knowledge gap in developing an actionable methodology for AI system development in manufacturing contexts.

Keywords: human-centred AI; manufacturing; AI system design; machine learning; socio-technical systems; design guidance.

Reference to this paper should be made as follows: Yamamoto, Y., Álvaro, A.M. and Sandström, K. (2024) 'Challenges in designing a human-centred AI system in manufacturing', *Int. J. Mechatronics and Manufacturing Systems*, Vol. 17, No. 4, pp.351–369.

Biographical notes: Yuji Yamamoto is an Assistant Professor at Mälardalen University, specialising in Operations Management and the Integration of Digital Technologies in Production. With a PhD in production management from Mälardalen University, he has extensive experience in teaching and research, focusing on the early phases of production system development, organisational capacity for innovation, and implementation of the Internet of Things and Artificial Intelligence from a socio-technical perspective.

Aranda Muñoz Álvaro is a Computer Scientist who holds a PhD in Innovation and Design. With over a decade of experience in research institutes, he has worked on topics of Human-Computer Interaction primarily in the context of industrial settings. His work underscores participatory and collaborative design, with the overall aim to support others in co-creating innovative and sustainable solutions with technologies like the Internet of Things and Artificial Intelligence. This approach entails the creation and facilitation of materials and workshop methods to help participants understand what is possible and shape solutions collaboratively.

Kristian Sandström is a Professor of Computer Science at Mälardalen University, specialising in Industrial AI and Digital Services. He serves as the head of the Industrial Systems research subject and leads the Industrial AI Systems research group. His research focuses on leveraging AI to drive industrial transformation and develop innovative digital services. He is leading major research initiatives, including the "Generative AI for Industrial Transformation" and "IndTech DS3" projects. He earned his PhD from KTH Royal Institute of Technology in 2002 and has since played a key role in advancing research and applications in industrial AI and digital service innovation.

This paper is a revised and expanded version of a paper entitled 'Practical aspects of designing a human-centred AI system in manufacturing' presented at *5th International Conference on Industry 4.0 and Smart Manufacturing (ISM 2023)*, Lisbon, 22–24 November, 2023.

1 Introduction

Manufacturing enterprises increasingly recognise the immense potential of incorporating AI-driven applications, referred to as AI systems, into their production processes. For example, machine learning (ML) models can be utilised to monitor and predict the condition of manufacturing equipment or identify abnormal patterns in manufacturing-related data. ML models can augment production planning capabilities by predicting future material demands, resource availability, and lead times with improved accuracy (Wang et al., 2021) and can assist or automate visual inspection of product and component quality (Chen et al., 2021; Lee, 2020).

Despite the recognised potential, adopting AI technology in manufacturing presents numerous challenges. The technology is often employed in high-stakes domains where flawed judgements, decisions, or controls can have severe detrimental effects on productivity, quality, speed, and safety in manufacturing operations (Kaymakci et al., 2021; Lee, 2020). Data availability and quality pose significant obstacles when developing and implementing AI models in manufacturing (Arinez et al., 2020). Constructing a high-performance ML model typically necessitates a substantial amount of structured data, which may be limited or challenging to obtain in practice. Furthermore, manufacturing data is often dynamic, exhibiting asynchronous changes due to variations in equipment, personnel, manufacturing processes, and materials. ML model prototypes are often created using data from specific periods of dynamic data (Kaymakci et al., 2021).

The significant knowledge gap between manufacturing personnel and data scientists is another major challenge (Arinez et al., 2020). Manufacturing professionals generally possess limited knowledge of ML techniques, making it difficult to draw realistic expectations or specifications for AI systems. Conversely, only a handful of data scientists possess profound insights into shop floor operations. This may result in AI system development disregarding domain experts' valuable knowledge, experience, and creativity (Makarius et al., 2020). Moreover, the potential impact of AI on existing jobs may generate fear and resistance among employees (Dabbous et al., 2022).

There is a strong need for a systematic approach to developing and integrating AI systems into manufacturing processes while addressing those challenges. However, methodological support for AI system development in manufacturing remains underdeveloped. Previous studies focus on the initial phases of the development lifecycle, such as problem analysis and ideation of AI systems (Emmanouilidis et al., 2021; Emmanouilidis and Waschull, 2021; Waschull and Emmanouilidis, 2022). Outside the manufacturing domain, scholars and practitioners have proposed AI system development process models and design recommendations for AI systems (Assadi et al., 2022; Google, 2023a; Subramonyam et al., 2022; van der Vegt et al., 2023). However, these guiding artefacts primarily stem from data science, user experience, and software engineering domains or are intended for non-manufacturing areas such as healthcare and consumer software products. It is thus uncertain how well they address manufacturing-specific contexts and challenges. Manufacturing operations are often regarded as socio-technical systems, where the symbiosis of production personnel and an AI system is considered critical for the implementation success (Emmanouilidis and Waschull, 2021; Makarius et al., 2020).

Our long-term research objective is to develop an actionable methodology for developing and integrating AI systems in manufacturing. This paper focuses on the early design phase of the AI system development life cycle, which involves understanding and integrating stakeholder desires, needs, and requirements into a system description using the domain language (Sommerville, 2016; Sundar, 2010). The paper also focuses on *guidance* for this phase, meaning clearly defined recommendations or procedures to achieve tasks in the phase. The study aims to identify shortages of existing guidance in the manufacturing literature.

A literature review and an embedded case were conducted to achieve the study's aim. In the case study, the researchers organised co-design workshops for two AI projects. The case study revealed the thinking-level challenges that design participants might face in the early design phase. They contribute to the design phase's complexity and inscrutability, and they may lead to cognitive overload among the design participants unless properly managed. The study also identified remedies for those challenges. These remedies, such as forming cross-functional teams for design activities, have already been recognised in the literature, but their association with those challenges has been unclear. The study concludes that the existing guidance in the manufacturing literature falls short of effectively addressing those challenges.

The remainder of this paper is structured as follows: The next section reviews the previous studies related to the main topic of the present study. Section 3 explains the case study method, and its results are presented in Section 4. Section 5 discusses the identified challenges and remedies and how they have been dealt with in the current literature. The last section draws conclusions and discusses the study's contributions.

2 Related research

This section reviews the literature related to this paper's main topic. It first conceptualises an AI system and discusses human involvement in AI system design. Then, it reviews existing guidance for the early design phase of AI system development in manufacturing research and other research domains.

2.1 AI systems in manufacturing

The construct of an AI system should be clarified. In the manufacturing literature, various conceptual models of an AI system have been proposed (Bousdekis et al., 2020; Emmanouilidis and Waschull, 2021; Kaymakci et al., 2021; Lee, 2020; Rožanec et al., 2022). In this paper, an AI system is understood based on the simple model suggested by Kaymakci et al. (2021). In this model, an AI system receives data from external sources as inputs and interacts with humans or external systems as outputs. According to the model, an AI system comprises at least three types of data processing units: data pre-processing units, ML models, and agents. The data pre-processing unit receives input data and processes them into a form that the ML model can use. The ML model takes in the data from the pre-processed data and generates inferences. The agent transforms the outputs from the ML model into a form that can interact with humans and other programs outside of the system. This model of an AI system is a technical system. In this paper, a solution system refers to a socio-technical system where humans interact with the technical system to provide solutions to manufacturing-related processes.

The study in this paper is interested in two aspects of human involvement in AI system design:

1 the participation of production personnel in the design

2 the establishment of human-AI symbiosis.

Regarding the first aspect, production personnel, including operators, supervisors, technicians and engineers, are often the primary users of AI systems in manufacturing contexts. The manufacturing literature recognises the significance of their participation because it enables the adequate utilisation of their creativity and expertise in design, leading to their empowerment (Hannola et al., 2018; Michael et al., 2019).

The second aspect is the establishment of human-AI symbiosis. In this concept, humans and AI effectively collaborate in manufacturing operations (Bousdekis et al., 2020; Emmanouilidis and Waschull, 2021). For instance, an AI system can aid factory employees in obtaining situational awareness or analytical insights for improved decision-making, while employees can provide feedback to refine the system's performance (Bousdekis et al., 2020). Although the desirable human-AI symbiosis in a manufacturing context has been envisioned in conceptual models (Bousdekis et al., 2020; Emmanouilidis and Waschull, 2021), the literature scarcely discusses how to practically design such symbiotic systems.

2.2 Guidance for the early design phase of AI system development in manufacturing

The software engineering literature describes that the design phase in the information system development lifecycle involves refining needs and requirements for the system and integrating them into a system description suitable for the implementation – coding with programming languages (Sommerville, 2016; Sundar, 2010). At the highest level of abstraction, a system description is an explanation of how the needs and requirements are technically addressed using broad terms and the language of the problem environment. At lower levels of abstraction, a more procedural language is used, such as UML (Sundar, 2010). This paper concentrates on an earlier design phase, which means refining and integrating needs and requirements into a system description in broad terms.

In the manufacturing literature, only a limited number of previous studies discuss and suggest guidance for the early design phase (Emmanouilidis et al., 2021; Emmanouilidis and Waschull, 2021; Ipektsidis and Soldatos, 2021a, 2021b, 2021c; Kaymakci et al., 2021; Waschull and Emmanouilidis, 2022). Kaymakci et al. (2021) present a process model of the AI system development life cycle containing four stages: planning, experimentation, implementation, and operation. The early design phase corresponds to the first stage, where the authors recommend that the design actor develop a first draft of an AI system, including its input data, functionality, system performance, and usage environment. However, the granularity of the recommendations remains at this level, which is too general for practitioners to turn into practice. Furthermore, the recommendations mostly concern software system design and not the user's and other stakeholders' participation in the design, nor the realisation of human-AI symbiotic systems.

The remaining references found in the manufacturing literature are from the same research project called the STAR (Emmanouilidis et al., 2021; Emmanouilidis and Waschull, 2021; Ipektsidis and Soldatos, 2021a, 2021b, 2021c; Waschull and Emmanouilidis, 2022). The project's objective is similar to our long-term research goal: to develop an actionable methodology for developing and implementing an AI system in manufacturing. The STAR project also highlights stakeholder participation and the realisation of human-AI symbiotic systems. This project identifies three phases of AI system development: definition and design, early development and testing, and final development and testing. The present study is relevant to the first phase, which includes three sub-steps: identifying the needs and requirements of the AI system, identifying the success criteria of the system, and then identifying technical components relevant to those needs and requirements. In the first two sub-steps, design participants are recommended to consider various issues related to human-AI symbioses, such as the systems' trustworthiness, safety, and explainability, and user feedback to the system. However, recommendations for the third step are unextensive. It only recommends that users and technology experts work together to identify physical entities associated with the needs and requirements, such as cameras, PLCs, and human-machine interface screens.

2.3 Guidance for the early AI system design in other research domains

The literature review in the present study was extended to other research areas due to the scarcity of relevant studies in the manufacturing domain. Two areas were identified: human-AI interaction design and AI system development in clinical research.

In the first area, researchers in academia and at leading software companies such as Google, Microsoft, and IBM have proposed various sets of recommendations that design actors should consider during the early design of AI systems (Google, 2023a; IBM, 2023; Microsoft, 2023; Piet, 2019). These sets of recommendations are often called Human-AI (HAI) design guidelines. Microsoft research proposes 18 design guidelines (Amershi et al., 2019), and The Google People + AI (PAIR) research team proposes 23 guidelines (Google, 2023a). These guidelines are grounded in design thinking, user experience (UX), and AI ethics theories and practices (Subramonyam et al., 2022). Examples of guidelines are 'Give control back to the user when automation fails' and 'Enable the user to provide feedback on their preferences during regular interaction with the AI system' (Google, 2023a). PAIR has published a design guidebook showing how to organise design workshops utilising those guidelines (Google, 2023a). The guidebook suggests three steps: introducing relevant AI capabilities to stakeholders, identifying problems that may occur during human-AI system interactions, and detailing mitigation plans addressing those problems by consulting with the guidelines.

The guidelines and the guidebook are relevant to the present study because they assist in the early design of human-AI symbiotic systems with stakeholder involvement. The recommendations in those guidelines are more detailed, concrete, and empirically grounded than those found in the manufacturing literature. However, their applicability to the manufacturing domain is uncertain. While the usefulness of these HAI guidelines has been tested by Amershi et al. (2019), the test was limited to designing AI-enabled consumer applications such as email clients, social network software products, and music players. These applications are for low-stake areas and may not be immediately relevant to manufacturing.

In the field of clinical research, several scholars have presented process frameworks for AI system development in clinical environments (Assadi et al., 2022; Gu et al., 2021; de Hond et al., 2022; Sendak et al., 2020; van der Vegt et al., 2023). The number of relevant publications in this research exceeds that of manufacturing research, implying that methodological development is more advanced in the former research.

AI-enabled clinical applications often aim at high-stakes areas, such as pathology diagnosis (Gu et al., 2023) and sepsis detection (Sendak et al., 2020). Applications are often intended to realise human-AI symbiosis, assisting clinicians' and nurses' effective decision-making. Seamlessly integrating the system into the clinical workflows is considered critical for successful implementations (Sandhu et al., 2020; Sendak et al., 2020). The high-stake applications and the need for human-AI symbiosis and workflow integration are relevant to manufacturing contexts.

The process frameworks contain varying numbers of phases. At the phases relevant to the early design, the frameworks include recommendations, such as 'updating problem descriptions and success criteria' (de Hond et al., 2022; Van De Sande et al., 2022), 'identifying relevant data and accounting for bias and privacy' (Assadi et al., 2022), and 'assessing and planning for risks and consequences of system errors' (de Hond et al., 2022).

The literature review indicates that the granularity of those recommendations lies between those in the manufacturing research and the HAI guidelines. The applicability of these recommendations to the manufacturing domain is unproven, as the literature review did not identify any study attempting to apply them to manufacturing contexts.

Overall, the literature review reveals that the existing guidance for the early design phase mostly concerns the questions of what features and functions of the human-AI symbiotic systems the design actors need to consider during the early design phase. Table 1 shows the categories of those features and functions found in the review. Those categories seem to be broad and reasonably comprehensive. Limitations in the guidance are also found. The limitations in the manufacturing literature are already mentioned in Section 2.2. The applicability of the guidance found in human-AI interaction and clinical research is unproven in the manufacturing domain.

Table 1 Categories AI system features that the three research areas recommend to consider in the early design

Categories of AI system features recommended to be considered in the early system design	*Guidance found in manufacturing research*	*Guidance found in the human-AI interaction research*	*Guidance found in clinical research*
Balance of control and automation	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
Presentation of inference results to users	x	x	x
Input data and its collection and labelling methods		x	x
Security, safety, and privacy	x	x	x
AI model and system performance	x	x	x
Integration into the operational workflows	x		x
Integration into the information system infrastructure	x		x
References	Emmanouilidis et al. (2021), Emmanouilidis and Waschull (2021), Ipektsidis and Soldatos (2021a, 2021b, 2021c), Kaymakci et al. (2021), Waschull and Emmanouilidis (2022)	Apple (2023), Google (2023a), IBM (2023), Microsoft (2023) and Piet (2019)	Assadi et al. (2022), Gu et al. (2021), de Hond et al. (2022), Van De Sande et al. (2022), Sandhu et al. (2020), Sendak et al. (2020) and van der Vegt et al. (2023)

3 Case study method

An embedded case study (Yin, 2011) was conducted at a manufacturing company in Sweden to further explore the limitations of the currently available guidance by applying it in real manufacturing settings.

The design science described by Holmström et al. (2009) was employed in the case study. This research approach was chosen for two reasons. First, the lack of structured support for the early design of an AI system in a manufacturing context was an ill-structured problem (Simon, 1973) for most industrial companies. Second, the approach enabled effective role separation and collaboration between industry and academia. In the case study, researchers played the role of developing and experimenting with guidance for early design, while case study participants were responsible for developing AI systems with methodological support from the researchers. Since the purpose of the case study is to understand the problem of current guidance, the case study corresponds to the first phase of Holmström et al.'s (2009) four-phase design science model: *solution incubation – framing the problem and developing rudimentary solutions*.

Design science, however, entails risks of researcher bias in interventions and of failing to create effective solutions for the case company due to their lack of contextual knowledge (Holmström et al., 2009). To mitigate these risks, the study was conducted by a multidisciplinary research team with industry experience. The team consisted of researchers with expertise in adopting new technology in manufacturing, industrial AI system development, and user experience. Their combined years of practical experience and close collaboration with practitioners at the case company created intersubjectivity, reducing the risk of researcher bias and solution creation failure.

The case company is a large company manufacturing power distribution equipment. The researchers participated in the early design phase of two AI system development projects. The first project, Project A, developed an early prototype of an anomaly detection ML model for a casting process. The model was intended to detect anomalies in temperature data from sensors installed in the furnaces. With that prototype, the project team wanted to design other parts of the AI system, particularly the agent part of the AI system.

The researcher intervened in the early design phase by organising and facilitating a design workshop with project members utilising the guidance in the literature. The researchers chose the HAI guidelines and the guidebook developed by PAIR (Google, 2023a) because of its concreteness, inclusion of workshop guidance, and a strong focus on participatory design. The workshop was conducted in February 2023 for 4 h with participants from R&D, process engineering, data science, manufacturing technicians, and Lean Six Sigma experts. The workshop aimed to generate a system concept considering the integration of the ML model into the manufacturing operations. As suggested in the PAIR guidebook, the workshop contained the following sessions: a warm-up session to familiarise participants with human-AI symbiotic systems, the initial design of user interfaces, and the identification of potential challenges in shopfloor integration and corresponding mitigation plans.

Another project, Project B, was to develop a real-time quality control system for an assembly operation using image recognition technologies. The context of the research intervention was similar to Project A, where the company had developed an early prototype of the ML model and wanted to understand how the model could be integrated into the operations. The researcher organised and facilitated a design workshop with the same purpose as the previous workshop conducted 10 months earlier. The workshop content was redesigned based on the learning from the first workshop. The workshop lasted 3 h and included participants from quality engineering, production engineering, R&D, and assembly operators. The workshop sessions included a warm-up session introducing AI vision technologies, the initial design of user interfaces, and the identification of potential challenges in shopfloor integration and mitigation plans.

During the workshops, the researchers observed and listened to participants' behaviours and conversations, which were audio recorded. Post-workshop reflections were held among the researchers and a few participants. The researchers were aware of the limited generality of a 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 (Westbrook, 1995).

4 Case study results

This section presents the results of the two design workshops held at the case company.

4.1 Design workshops for project A

The first session of the workshop was to introduce key aspects of designing a human-AI symbiotic system to the workshop participants to increase their understanding of the subject.

In the second session, the participants discussed two topics: how the AI system should be integrated into operational procedures (i.e., workflows) on the shop floor and how the system should present the ML model outputs to the users. For the first topic, the general consensus was that the furnace workers would be the first to receive the alarm from the AI system. They would check the system's user interface screen and decide on further actions, which might involve diagnosing the alarm themselves or consulting process engineers. If necessary, the engineers could contact equipment suppliers for further diagnosis.

For the second topic, the participants discussed the interface design for the screen. Various ideas were generated. For instance, the screen should provide a comprehensive shop floor overview of the furnace section, highlighting the specific furnace where anomalous signals were detected. Additionally, it should display detailed information about the anomaly, including its underlying time-series data and the historical actions taken in response to previous alarms.

While various ideas for the workflows and user interface design were generated in this session, the participants struggled to detail the design further. They experienced that the workflow and interface design were more complex than expected. An anomaly detection alarm would cause multiple branches of workflows. The branching would depend on various factors, for instance, the skills and knowledge of the operators and engineers (e.g., senior operators might be able to diagnose the alarm, but others would consult with engineers) and the mode, urgency, and confidence score of the anomaly (e.g., operators might need to react an alarm immediately regardless of its confidence score if it is urgent). The participants realised that user interface design should reflect those manifold scenarios, which they found not straightforward.

The third session focused on identifying potential challenges in the interaction between humans and AI systems and generating mitigation plans. To facilitate this activity, the participants consulted the 23 design guidelines from the PAIR guidebook. Various challenges were discussed, including alarm fatigue due to too many false positives, uncertainty in detecting false negatives, and the maintenance of the ML model due to data drift from the sensors. To mitigate alarm fatigue, it was suggested that the project team should communicate carefully with furnace operators about the potential evolution of system performance. To address maintenance, regular meetings with data scientists and production personnel were proposed to discuss model maintenance issues.

Although this session encouraged the participants to explore various challenges and corrective plans, they felt it was even more mentally exhausting than the previous session, as most discussions were based on imagination. They had to imagine possible problems based on hypothetical scenarios and further imagined solutions without exactly knowing the final behaviours of the system.

Overall, the workshop was not particularly successful in that they felt had to deal with too many design-related questions simultaneously, without clear navigation of how to answer those questions in which order. They experienced that questions kept diverging without converging. One participant's comment summarises this sentiment: "*If you compare this* (*developing an AI system*) *to building a house, in the pre-study, we put up wallpaper* (*an ML model*)*, and it looked interesting, but we haven't seen the design of the whole house, …and it is much more complex than we thought. We need a better plan to build and integrate all these different parts in the right order*".

4.2 Design workshop for project B

4.2.1 Improvements based on the learning from project A

The design workshop for Project B incorporated several improvements based on the insights gained from the previous workshop for Project A. The following three modifications were implemented.

The first one was the inclusion of operators in the workshop. During the Project A workshop, operators, who were the primary system users, were not present. Their absence led to increased guesswork in the design of workflows and interfaces, as other domain experts, such as process technicians and Lean Six Sigma experts, could not provide detailed operational insights. The prime system users in Project B were assembly operators, and they were included in the workshop.

The second modification was the controlled use of design guidelines. In the previous workshop, 23 design guidelines from PAIR were exposed to all participants to promote transparency. However, this approach confused the participants regarding which questions should be dealt with in which order. For the second workshop, the researchers selected and compiled a list of a limited number of questions from various sources, including the PAIR's design guidelines, as well as other HAI guidelines (IBM, 2023; Microsoft, 2023; Piet, 2019), and guidance from the clinical research (Assadi et al., 2022; de Hond et al., 2022; van der Vegt et al., 2023). The workshop facilitator held the list and posed questions based on the discussion's progress.

The third improvement was the use of boundary objects in the workshop. Boundary objects are artefacts that coordinate perspectives and align disparate communities of practice, often temporarily, to solve specific design problems (Beddoes and Nicewonger, 2019). They provide a shared syntax or language that individuals can use to represent their ideas, concepts, and knowledge (Broberg, 2011). They are often employed in participatory design practices (Broberg, 2011). As suggested by these authors, boundary objects could reduce the guesswork and increase the shared understanding among the participants through tangible design representations. The researcher used a business origami toolkit developed by Munoz et al. (2024) and a printed picture taken from the AI vision system as boundary objects for the workflow and interface design. The toolkit was chosen due to its proven usefulness in designing Industry 4.0 solutions in manufacturing environments.

4.2.2 Results of the three sessions in the workshop

The first workshop session introduced participants to AI vision technologies using a Teachable Machine (Google, 2023b). Participants trained and tested an AI vision model, allowing them to understand its functionality and probabilistic behaviours and grasp the potential and limitations of the vision technologies.

The second session focused on designing the user interface for the intended AI system: a real-time assembly quality control system. The participants used the aforementioned camera image as a basis for discussion and generated several design ideas. For instance, quality control results should be displayed on the current assembly instruction screen while keeping the control result information minimal unless deviations would be detected. Additionally, the same screen should allow users to review previous control results in the assembly sequence to ensure the correct execution of the assembly process.

The third session concentrated on integrating the AI system into the operators' workflows. The participants explored and demonstrated various user scenarios using the business origami toolkit. For instance, low confidence scores in quality control results would trigger caution alarms, prompting operators to manually assess the quality and provide feedback to the system. Junior operators could consult with mentors for assessment assistance. As another example, the participants discussed whether quality deviation alarms should be displayed on mobile devices. It was deemed unnecessary, as operators typically work from a fixed position during assembly. A dedicated fixed-position screen monitor would suffice for their user interaction needs.

During the second and third sessions, the workshop facilitators referred to the aforementioned question list and posed questions to ensure that important design concerns were addressed. Examples of the questions posed included 'How many false positives per day can the operator tolerate?' and 'What misunderstandings may arise when introducing the system to other operators?'.

Overall, the workshop exhibited better organisation compared to the previous one. The researchers received no feedback from the participants expressing confusion or cognitive overload. The three improvements based on the learnings from the first workshop seemed to yield a positive effect. The involvement of the operators made the discussions more concrete and concise. The boundary objects used in the workshop served as central points for discussions and contributed to the shared understanding of multiple user scenarios. The controlled use of design questions prevented the discussions from diverging and maintained a clear focus on the discussion topics.

5 Cognitive challenges and their remedies in the early design phase

It is recognisable in the case study that the researchers and industry participants struggled to deal with complex and multifaceted aspects of the early design of human-AI symbiotic systems to reduce the inherent fuzziness and ambiguity of this phase. An analysis of the case study has revealed five cognitive challenges that added complexity and uncertainty to the design process, increasing the mental effort and work memory required of design participants. The analysis has also led to identifying five remedies related to those cognitive challenges. The conceptual cause-and-effect relationship among those cognitive challenges, design participants' mental stains, and remedies is depicted in Figure 1. The five challenges and remedies are summarised in Table 2.

Figure 1 Conceptual relationship among cognitive challenges, remedies, and mental strain on design participants

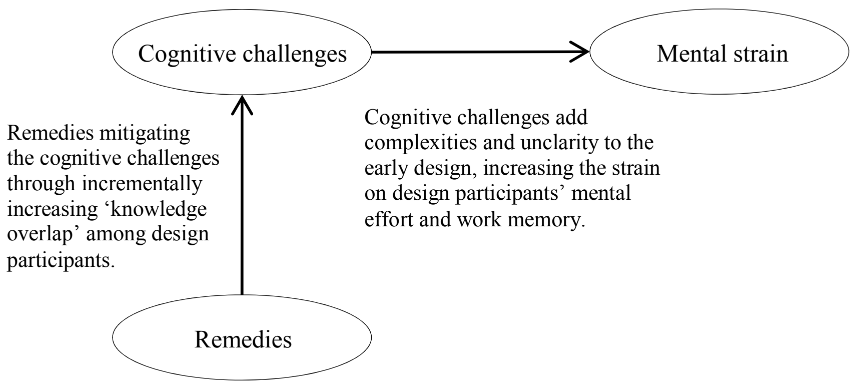

Table 2 Cognitive challenges and remedies found in the study

Cognitive challenges	*Remedies*
C1: Complex patterns of operational workflows integrating an AI system	R1: Forming a cross-functional team and ensuring the presence of key persons at design activities
C2: Uncertainty about data and ML model outputs	R2: Introducing AI capabilities and limitations to non-experts
C3: Interdependency of design components	R3: Use of boundary objects (e.g., swimlanes, business origamis, prototypes, demos)
C4: Absence of the holistic scope	R4: Adopting participatory design practices
C5: Difficulties in prioritising inquiries	R5: Incremental delivery of design results through controlled use of design guidelines

The first part of this section describes the cognitive challenges, and the second part discusses the five remedies found. Based on these findings, the third subsection discusses the shortage of existing guidance in the manufacturing literature for the early design phase of AI system development.

5.1 Five cognitive challenges

Challenge 1: *Complex patterns of operational workflows integrating an AI system:* As indicated in the first workshop in the case study, designing operational workflows that effectively integrate an AI system can be a complex task, at least in some use cases. The first workshop demonstrated that an anomaly detection alarm could trigger branching patterns of operational procedures involving multiple actors with diverse professions. The branching patterns were found to be contingent on various factors, such as individual skills and knowledge of operators and engineers, their availability at workplaces, and the urgency and confidence scores of the alarm. Moreover, other events could initiate other branching workflows, for instance, when operators suspect false negatives or degradation of model inference due to data shift.

Challenge 2: *Uncertainty about data and ML model outputs:* The literature recognises this challenge (Kaymakci et al., 2021; Lee, 2020), and it was also evident in the case study. Outputs from an ML model inherit uncertainties due to its probability-based algorithms. In manufacturing contexts, the input data to the model are often dynamic, shifting over time post-deployment, which amplifies output uncertainty. It was observed in the case study that designing workflows, which typically adhere to rule-based logic, is instrumental in absorbing those uncertainties and preventing them from adversely affecting the manufacturing processes that usually aim for high precision and low variance.

Challenge 3: *Interdependency of design components:* van der Vegt et al. (2023), in clinical research, suggest that the following four components should be considered in designing an AI-integrated solution system: workflow, user interface, Al model, and data. The case study found that the properties and design of those components were significantly interdependent. For instance, during the workshops, it was observed that the user interface design was heavily dependent on the workflow design. The participants also noted that the interface design would depend on the ML model functionality and performance, which in turn would be affected by the input data and pre-process methods. This interdependence adds complexity to the early design phase and can easily lead to cognitive overload among design participants.

Challenge 4: Absence of the holistic scope

Different professionals are involved in the design of the aforementioned four components. According to the literature, data engineers and data scientists often focus on the data and AI model components (Kaymakci et al., 2021; Lee, 2020), user experience designers on the user interface component (Subramonyam et al., 2022), and users or domain experts on the workflow (Kutz et al., 2023). However, the case study evidenced that even though all necessary professionals were present in the workshops, no one had a clear and holistic picture of the solution system. Each professional could only see one aspect of the solution system based on their expertise and the immediately relevant components. This lack of a holistic perspective, combined with the component interdependency, substantially adds ambiguity to the early design.

Challenge 5: Difficulty in prioritising inquiries

In the first workshop, all of the 23 PAIR guidelines were exposed to the participants. This caused their experience of dealing with too many design questions simultaneously without clear guidance on prioritisation. The researchers learned from this failure, and in the second workshop, they prepared a limited number of questions and arbitrarily used them based on the progress of the discussion. This led to a more organised workshop and less divergent discussion. However, the researchers still felt uncertain about the prioritisation, appropriateness, and sufficiency of the questions posed.

5.2 Five remedies for the cognitive challenges

The five remedies identified in this study collectively contribute to incrementally increasing knowledge overlap among professionals involved in early AI system design, thereby reducing uncertainties and complexities tied to the five cognitive challenges. In this context, knowledge overlap refers to the ability of design participants to comprehend the contents and properties of other solution system components beyond their immediately relevant ones. Each remedy addresses multiple cognitive challenges, as shown in Table 3. Our analysis shows that these remedies were employed in at least one of the workshops conducted at the case company. This is also shown in Table 3.

Table 3 The relationship of the five remedies (R1-5) to the five cognitive challenges (C1-5). The relationship is made based on the authors' assessment. The table also shows in which workshop (WS) in the case study the remedies were employed

	R1	*R2*	*R3*	*R4*	*R5*
C1	x		x	x	
C2	x	x	x	x	
C3	x	x	x	x	x
C4	x	x	x	x	x
C5					x
Remedies employed	WS1 WS2	WS1 WS2	WS2	WS1 WS2	WS2

While these remedies have been acknowledged in existing literature, previous studies have not linked them to the cognitive challenges with the same level of detail as the current study. For instance, in the reviewed manufacturing literature, Emmanouilidis and Waschull (2021) and Waschull and Emmanouilidis (2022) highlight the significance of forming cross-functional teams (R1 in Table 2) and adopting participatory design (R4) to enhance collaboration among design participants. However, these authors do not specifically link these remedies to the cognitive challenges.

The human-AI interaction research and the clinic research discuss the five remedies and their connection to one of a few cognitive challenges. For instance, Subramonyam et al. (2022) suggest using boundary objects (R3), such as program scripts and user interface prototypes, to address the interdependency between AI model development and interface design. The interdependency corresponds to C3 but only between the AI model and interface components. In clinical research, several researchers advocate for using swimlanes to design clinical workflows integrating AI systems (Gu et al., 2023; Sandhu et al., 2020). The use of swimlanes is discussed without explicitly relating to the cognitive challenges, but it can be assumed that the swimlanes are boundary objects facilitating the simultaneous design of workflow and user interface components, which mitigates the challenge of component interdependency (C3). The swimlanes can also help to make potentially complex workflow designs more visually comprehensible, addressing C1.

The HAI guidelines (Google, 2023a; Piet, 2019) suggest that introducing relevant AI technologies to design participants can increase their comprehension of the technology's capabilities and limitations (R2). Microsoft (2023) suggest selecting specific design guidelines out of their 18 guidelines before a design activity, which is relevant to R5. These sources, however, discuss these remedies without explicit ties to any of the cognitive challenges.

The remedies were employed in one or both of the workshops, as shown in Table 3. R1, R2, and R4 were employed in the first workshop, but participants still reported experiencing cognitive overload. The second workshop incorporated all five remedies, resulting in a significantly more organised and less confusing design activity. This observation may imply that the five cognitive challenges can be approached by broadly employing those remedies rather than selecting a few of them.

5.3 Shortage in the guidance

The analysis presented in this section enables us to specify the shortage in the existing guidance in the manufacturing literature for the early design phase of AI system development, which is the purpose of the present study. The study has focused on the design of human-AI symbiotic systems.

The identified shortage can be summarised as follows: While the existing guidance in manufacturing literature has advanced in outlining what system features and functions, such as those categorised in Table 1, design participants need to consider during the early design phase of developing AI systems, it falls short in providing detailed procedural knowledge on how to conduct design activities, taking those cognitive challenges into account and effectively incorporating those remedies in the design phase.

The current guidance merely provides a basic procedure for the early phase, such as the three-substep procedure described in Section 2.2. The manufacturing literature and other reviewed literature have scarcely discussed those cognitive challenges, at least not to the depth explored in this study. It appears that in previous studies, the researchers and practitioners have arbitrarily employed one or few of those remedies based on their prior experience without a strong awareness of effectively addressing those challenges.

We argue that future development of methodological support for the early phase should consider this identified shortage to adequately manage the inscrutability and complexity inherent in the design at this phase.

6 Conclusions and discussions

The purpose of the study presented in this paper is to identify the shortage in the existing guidance within the manufacturing literature for the early design phase of developing AI systems. To this end, the literature and an embedded case study were conducted. The results revealed the five cognitive challenges faced by design participants in the early design phase, along with the five associated remedies. The study highlighted the complexity and inscrutability of this phase and concluded that the identified challenges are scarcely addressed and that the remedies are insufficiently incorporated in the current guidance.

The primary theoretical contribution of the present study is identifying the cognitive-level challenges faced by design participants during the early design phase of the AI system development in the manufacturing context. These cognitive challenges, if not properly managed, contribute to the complexity and inscrutability of this phase. The study found that existing research fails to provide sufficient knowledge that addresses these cognitive challenges effectively.

Sections 2 and 5.3 highlight that the previous studies in manufacturing research have primarily focused on what system features and functions design participants should consider in the early phase of developing a human-AI symbiotic system (Emmanouilidis and Waschull, 2021; Ipektsidis and Soldatos, 2021c; Kaymakci et al., 2021). This addresses 'what' questions. Interestingly, research on human-AI interaction also reveals a dearth of studies probing the 'how' question. Subramonyam et al. (2022) argue that the current HAI guidelines (Google, 2023a; IBM, 2023; Microsoft, 2023; Piet, 2019) mostly focus on 'what' needs to be done in design activities, for instance, "consider gaining users' trust in the AI system", but they make no recommendations about 'how' specific design processes serve to bridge knowledge boundaries across various professions.

The present study identified the five remedies to those cognitive challenges. While these antidotes have been known in the literature, their association with the challenges has not been unclear. This study contributes to the literature by providing a deeper understanding of why these remedies, such as using boundary objects, should be employed during the early design phase of AI system development in manufacturing.

A practical contribution of this study is that it demonstrates the difficulties of adopting the HAI guidelines (Google, 2023a; IBM, 2023; Microsoft, 2023; Piet, 2019) in manufacturing settings, especially when an AI system is applied to high-stakes domains. The case study evidenced that the PAIR guidelines were hardly applicable as they were, primarily because they did not effectively assist workflow design. As discussed in Section 5, the workflow design is critical to absorb the uncertainties of outputs from the AI models. This shortage seems to be attributed to the fact that the guidelines are primarily tailored to and draw knowledge from applications in relatively lower-stakes commercial application domains, such as social media applications and music players. The formulation of the second workshop in the case study can be considered another practical contribution, where the five remedies were taken into account, and the workshop did not lead to the design participants' cognitive overload. However, the generality of the formulation is questionable, as it was for a specific use case and mostly focused on the interface and workflow components of the solution system.

Finally, a key limitation of the present study and future research can be mentioned. The findings in this study are based on a single case study. Their validity and generalisability must be examined with multiple case studies. Considering the immaturity of the research area, more cognitive challenges and remedies may emerge in future research. The present study focused only on the early design phase. Understanding potential cognitive challenges and remedies in other phases of AI system development is another pathway for future research.

Acknowledgements

This research is funded by Vinnova under the project Democratisation of AI in manufacturing organisations with the reference number 2022-0128.

References

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P.N., Inkpen, K., Teevan, J., Kikin-Gil, R., Horvitz, E. and Assoc Comp, M. (2019) 'Guidelines for human-AI interaction', *Conference on Human Factors in Computing Systems – Proceedings*, doi: 10.1145/3290605.3300233.

Arinez, J.F., Chang, Q., Gao, R.X., Xu, C. and Zhang, J. (2020) 'Artificial intelligence in advanced manufacturing: current status and future outlook', *Journal of Manufacturing Science and Engineering*, *Transactions of the ASME*, Vol. 142, No. 11, pp.110804-1–110804-16, doi: 10.1115/1.4047855.

Assadi, A., Laussen, P.C., Goodwin, A.J., Goodfellow, S., Dixon, W., Greer, R.W., Jegatheeswaran, A., Singh, D., McCradden, M., Gallant, S.N., Goldenberg, A., Eytan, D. and Mazwi, M.L. (2022) 'An integration engineering framework for machine learning in healthcare', *Frontiers in Digital Health*, Frontiers, Vol. 4, p.932411.

Beddoes, K., and Nicewonger, T.E. (2019) 'Boundary negotiating artifacts for design communication: a theoretical and empirical exploration', *Education, Design and Practice – Understanding Skills in a Complex World*, https://par.nsf.gov/biblio/10137273

Bousdekis, A., Apostolou, D. and Mentzas, G. (2020) 'A human cyber physical system framework for operator 4.0 – artificial intelligence symbiosis', *Manufacturing Letters*, Vol. 25, pp.10–15, https://doi.org/10.1016/j.mfglet.2020.06.001

Broberg, O. (2011) 'Enabling objects for participatory design of socio-technical systems', *DS 68-7: Proceedings of the 18th International Conference on Engineering Design (ICED 11), Impacting Society through Engineering Design, Vol. 7: Human Behaviour in Design*, Lyngby/Copenhagen, 15–19 August, Denmark, pp.64–73.

Chen, J., Zhang, Z. and Wu, F. (2021) 'A data-driven method for enhancing the image-based automatic inspection of IC wire bonding defects', *International Journal of Production Research*, Vol. 59, No. 16, pp.4779–4793, doi: 10.1080/00207543.2020.1821928.

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

de Hond, A.A.H., Leeuwenberg, A.M., Hooft, L., Kant, I.M.J., Nijman, S.W.J., van Os, H.J.A., Aardoom, J.J., Debray, T.P.A., Schuit, E., van Smeden, M., Reitsma, J.B., Steyerberg, E.W., Chavannes, N.H. and Moons, K.G.M. (2022) 'Guidelines and quality criteria for artificial intelligence based prediction models in healthcare: a scoping review', *NPJ Digital Medicine*, *Med*, Vol. 5, No. 1, p.2.

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

Emmanouilidis, C., Waschull, S., Bokhorst, J.A.C. and Wortmann, J.C. (2021) 'Human in the AI loop in production environments', *IFIP Advances in Information and Communication Technology*, Vol. 633 IFIP, doi: 10.1007/978-3-030-85910-7_35.

Google (2023a) *People + AI Guidebook*, Available at: https://pair.withgoogle.com/guidebook (Accessed 24 June, 2023).

Google (2023b) *Teachable Machine*, Available at. https://teachablemachine.withgoogle.com/train (Accessed 15 December, 2023).

Gu, H., Huang, J., Hung, L. and Chen, X. (2021) 'Lessons learned from designing an AI-enabled diagnosis tool for pathologists', *Proceedings of the ACM on Human-Computer Interaction*, ACM, New York, NY, USA, Vol. 5, No. CSCW1, pp.1–25.

Gu, H.Y., Liang, Y., Xu, Y.F., Williams, C.K., Magaki, S., Khanlou, N., Vinters, H., Chen, Z.S., Ni, S., Yang, C.X., Yan, W.Z., Zhang, X.R., Li, Y., Haeri, M. and Chen, X. (2023) 'Improving workflow integration with XPath: Design and evaluation of a human-AI diagnosis system in pathology', *ACM Transactions on Computer-Human Interaction*, ACM New York, NY, Vol. 30, No. 2, pp.1–37.

Hannola, L., Richter, A., Richter, S. and Stocker, A. (2018) 'Empowering production workers with digitally facilitated knowledge processes – a conceptual framework', *International Journal of Production Research*, Taylor & Francis, [London], Vol. 56, No. 14, pp.4729–4743, doi: 10.1080/00207543.2018.1445877.

Holmström, J., Ketokivi, M. and Hameri, A.P. (2009) 'Bridging practice and theory: a design science approach', *Decision Sciences*, Vol. 40, No. 1, pp.65–87, doi: 10.1111/j.1540-5915.2008.00221.x.

IBM (2023) *AI Essentials Framework*, Available at. https://www.ibm.com/design/thinking/page/courses/AI_Essentials (Accessed 25 June, 2023).

Ipektsidis, C. and Soldatos, J. (2021a) *Reference Scenarios and Use Cases for AI in Manufacturing*, Available at: https://star-ai.eu/deliverables (Accessed 1 November, 2024).

Ipektsidis, C. and Soldatos, J. (2021b) *User Studies and Co-Design for Human-Centred AI*, Available at: https://star-ai.eu/deliverables (Accessed 1 November, 2024).

Ipektsidis, C. and Soldatos, J. (2021c) *Report on Co-Design Workshops and Focus Groups-Initial Version*, Available at: https://star-ai.eu/deliverables (Accessed 1 November, 2024).

Kaymakci, C., Wenninger, S. and Sauer, A. (2021) 'A holistic framework for AI systems in industrial applications', *Lecture Notes in Information Systems and Organisation*, Vol. 47, pp.78–93, doi: 10.1007/978-3-030-86797-3_6.

Kutz, J., Neuhüttler, J., Schaefer, K., Spilski, J. and Lachmann, T. (2023) 'Generic role model for the systematic development of internal AI-based services in manufacturing', *The 56th Hawaii International Conference on System Sciences*, Vol. 4, pp.909–917.

Lee, J. (2020) *Industrial AI: Applications with Sustainable Performance*, Springer, Singapore.

Makarius, E.E., Mukherjee, D., Fox, J.D. and Fox, A.K. (2020) 'Rising with the machines: a sociotechnical framework for bringing artificial intelligence into the organization', *Journal of Business Research*, Vol. 120, pp.262–273, doi: 10.1016/j. jbusres.2020.07.045.

Michael, L., Alexander, R. and Melanie, S. (2019) 'Power to the worker's: empowering shop floor workers with worker-centric digital designs', *International Journal of Operations and Production Management*, Emerald Publishing Limited, Vol. 39, No. 1, pp.24–42, doi: 10.1108/IJOPM-05-2017-0294.

Microsoft (2023) *The HAX Toolkit*, Available at. https://www.microsoft.com/en-us/haxtoolkit/ (Accessed 25 June, 2023).

Munoz, A.A., Yamamoto, Y., Sandström, K., Eriksson, Y. and Florin, U. (2024) 'The Karakuri IoT toolkit: a collaborative solution for ideating and prototyping IoT opportunities', *International Design Conference – DESIGN2024*, Dubrovnik, Croatia, pp.185–194.

Piet, N. (2019) *AI MEETS DESIGN TOOLKIT*, Available at. https://aixdesign.co/toolkit (Accessed 25 June, 2023).

Rožanec, J.M., Novalija, I., Zajec, P., Kenda, K., Tavakoli Ghinani, H., Suh, S., Veliou, E., Papamartzivanos, D., Giannetsos, T., Menesidou, S.A., Alonso, R., Cauli, N., Meloni, A., Recupero, D.R., Kyriazis, D., Sofianidis, G., Theodoropoulos, S., Fortuna, B., Mladenić, D. and Soldatos, J. (2022) 'Human-centric artificial intelligence architecture for industry 5.0 applications', *International Journal of Production Research*, doi: 10.1080/00207543.2022.2138611, Vol. 61, No. 20, pp.6847–6872.

Sandhu, S., Lin, A.L., Brajer, N., Sperling, J., Ratliff, W., Bedoya, A.D., Balu, S., O'Brien, C. and Sendak, M.P. (2020) 'Integrating a machine learning system into clinical workflows: qualitative study', *Journal of Medical Internet Research*, JMIR Publications Toronto, Canada, Vol. 22, No. 11, p.e22421.

Sendak, M.P., Ratliff, W., Sarro, D., Alderton, E., Futoma, J., Gao, M., Nichols, M., Revoir, M., Yashar, F., Miller, C., Kester, K., Sandhu, S., Corey, K., Brajer, N., Tan, C., Lin, A., Brown, T., Engelbosch, S., Anstrom, K., Elish, M.C., Heller, K., Donohoe, R., Theiling, J., Poon, E., Balu, S., Bedoya, A. and O'Brien, C. (2020) 'Real-world integration of a sepsis deep learning technology into routine clinical care: implementation study', *JMIR Medical Informatics*, Vol. 8, No. 7, doi: 10.2196/15182.

Simon, H.A. (1973) 'Does scientific discovery have a logic?', *Philosophy of Science*, Vol. 40, No. 4, pp.471–480, doi: 10.1086/288559.

Sommerville, I. (2016) *Software Engineering*, *Software Engineering*, 10th ed., Pearson, Boston.

Subramonyam, H., Im, J., Seifert, C. and Adar, E. (2022) *Human-AI Guidelines in Practice: Leaky Abstractions as an Enabler in Collaborative Software Teams*, arXiv preprint arXiv:2207.01749.

Sundar, D. (2010) *Software Engineering*, *Software Engineering*, 1st ed., University Science Press, an Imprint of Laxmi Publications Pvt. Ltd., New Delhi, India.

van de Sande, D., Van Genderen, M.E., Smit, J.M., Huiskens, J., Visser, J.J., Veen, R.E.R., van Unen, E., Ba, O.H., Gommers, D. and van Bommel, J. (2022) 'Developing, implementing and governing artificial intelligence in medicine: a step-by-step approach to prevent an artificial intelligence winter', *BMJ Health and Care Informatics*, BMJ Publishing Group, Vol. 29, No. 1, pp.1–8.

van der Vegt, A.H., Scott, I.A., Dermawan, K., Schnetler, R.J., Kalke, V.R. and Lane, P.J. (2023) 'Implementation frameworks for end-to-end clinical AI: derivation of the SALIENT framework', *Journal of the American Medical Informatics Association*, Oxford University Press, Vol. 30, No. 9, pp.1503–1515.

Wang, L., Pan, Z. and Wang, J. (2021) 'A review of reinforcement learning based intelligent optimization for manufacturing scheduling', *Complex System Modeling and Simulation*, Vol. 1, No. 4, pp.257–270, doi: 10.23919/CSMS.2021.0027.

Waschull, S. and Emmanouilidis, C. (2022) 'Development and application of a human-centric co-creation design method for AI-enabled systems in manufacturing', *IFAC-PapersOnLine*, Vol. 55, pp.516–521, doi: 10.1016/j. ifacol.2022.04.246.

Westbrook, R. (1995) 'Action research: a new paradigm for research in production and operations management', *International Journal of Operations and Production Management*, Vol. 15, No. 12, pp.6–20.

Yin, R.K. (2011) *Qualitative Research from Start to Finish*, Guilford Press, New York, NY.