6-18-2022

DESIGN PRINCIPLES FOR HUMAN-CENTRED AI

Stefan Cronholm
School of Business and IT, stefan.cronholm@hb.se

Hannes Göbel
University of Borås, hannes.gobel@hb.se

DESIGN PRINCIPLES FOR HUMAN-CENTRED AI

Research Paper

Stefan Cronholm, University of Borås, Borås, Sweden, stefan.cronholm@hb.se

Hannes Göbel, University of Borås, Borås, Sweden, hannes.gobel@hb.se

Abstract

Advancements within artificial intelligence (AI) enable organisations to reformulate strategies for exploiting data in order to refine their business models, make better decisions and maintain a competitive advantage. We recognise the technical advantages of AI. However, our view is that the technical perspective as a base for decision-making is necessary but insufficient. Several studies in human science report that essential human knowledge and competencies that affect decision making are not represented in AI systems. Based on this observation, we have developed design principles for developing decision-support systems (DSS) that combine human intelligence (HI) with AI. The design principles are: design for amplified decision-making, design for unbiased decision-making and design for human and AI learning. The design principles constitute the scientific contribution to the emergent field of Human-Centred AI. The contribution to practice consists of a DSS (a digital prototype) that supports the combination of HI and AI.

Keywords: Human-centred AI, Design principles, Decision-support systems, Action Design Research.

1 Introduction

The continuous development of artificial intelligence (AI) and advanced data analysis affects organisations in many ways. For example, advancements within AI enable organisations to reformulate strategies for exploiting data to refine business models, make better decisions and maintain competitive advantages (e.g., Russel and Norvig, 2016). The exploitation of data usually involves descriptive, predictive or prescriptive data analysis (Evans and Lindner, 2012) and machine learning (Jordan and Mitchell 2015; Meng et al. 2016). One purpose of machine learning is to automatically extracting predictive models from existing data in order to support decision making. Jordan and Mitchell (2015, p.255) state that: "The adoption of data-intensive machine-learning methods can be found throughout science, technology and commerce, leading to more evidence-based decision-making across many walks of life, including health care, manufacturing, education, financial modeling, policing, and marketing".

Furthermore, the advancements within AI spur organisations to generate branch-specific data (Cui et al., 2020). The growth of branch-specific data has implied that several organisations regard data as the most critical asset for making well-informed decisions (Demartini 2015; OECD, 2015; Cronholm et al., 2017). Doubtless, organisations are showing an increased interest in the development of AI and the exploitation of data. However, with the development of AI technology, the AI community has begun to realise that intelligent machines cannot completely replace human intelligence (e.g., intuition, consciousness, reasoning, abstract thinking) (Xu and Dainoff, 2021). As a consequence, concepts such as explainable AI, responsible AI, trustworthy AI and ethical AI are being frequently researched. Moreover, followers of the growing scientific field of Human-Centred AI (HCAI) argue that relying entirely on AI for decision making might be too risky (e.g., Xu, 2019; Shneiderman, 2021a). HCAI is an emerging field that aims to synthesise AI algorithms with human thinking (Shneiderman, 2021b). Moreover, HCAI stipulates that learning occurs from human input, machine-based analysis and collaboration between the two.

On the contrary, several studies within the area of human science report that essential human knowledge and competencies that affect decision making are not represented in AI systems (e.g., Kahneman 2011;

Demartini 2015; Borst 2016). Such knowledge and competencies are not based on technical algorithms or mathematical calculations of large data volumes. Instead, they are stored in organisational structures, personal memories and cognitive thought patterns (Göranzon, 2009). Furthermore, such knowledge is often based on professional experiences, personal reflections, branch-specific events, contextual factors, relationships to other involved actors, and organisational culture (ibid.). Kahneman (2011) states that the competence that AI systems lack is the human specific capability to apply intelligent cognitive processes, that is, the use of the brain. Finally, Shneiderman (2021, p.57) adds that "… standard approaches to machine learning do not allow adequate learning from such apparently exceptional events. Efforts to develop common-sense reasoning, explainable AI, and causal understanding seem still shallow compared with what humans do when they formulate problems, find innovative solutions, and—as I am doing here—raise challenges to existing beliefs". To simplify, we refer to human knowledge and experience as human intelligence (HI).

We acknowledge the advancements within AI, however, as stated above, human knowledge and experiences are in many ways superior to AI. Against this backdrop, our view is that a technical perspective is necessary but insufficient. Consequently, we argue that a social-technical perspective is important to apply in decision-making situations. Mumford (2006, p.321) defines a socio-technical perspective as "the joint optimization of the social and technical systems". The idea is that the relationships between socio and technical elements will increase productivity and make better decisions. From a socio-technical perspective, it seems as a good idea to combine human knowledge and experience with AI in order to achieve a stronger information basis for decision-making. Our assumption is strengthened by recent work from researchers stating that the integration of HI and AI intelligence provides a significant potential of augmentation for decision making (Zheng et al., 2017; Dellermann et al., 2019; Johnson and Vera, 2019). We define decision making as a process of making choices that reduces cognitive strain. It involves the selection of the best option from a choice set containing two or more options (Beach, 1993)

We have identified several examples of decision-support systems (DSS) on the market that support automatic decision making (Capterra, 2021). However, to our knowledge, there are no DSS that include the combination of HI and AI. Moreover, we have not found methods, principles or guidelines supporting the development of such tools. The problem we have formulated reads: there is a lack of design knowledge concerning the development of DSS, based on a combination of HI and AI. Our research question reads: How can DSS be designed through a combination of HI and AI? The scientific contribution consists of design knowledge expressed as design principles. Design principles constitute a prescriptive component included in design theory (Chandra et al., 2015). Moreover, the design principles constitute an important vehicle "… to convey design knowledge that contributes beyond instantiations applicable in a limited use context is that of a *design principle"* (ibid., p.4039). The design knowledge developed in our study is based on theoretical insights and empirical evidence from evaluating the DSS in the retail sector. We view the design knowledge as a contribution to the field of HCAI (see section 2) while the contribution to practice consists of a DSS (digital prototype) that supports decision making based on the combination of HI and AI.

We limited our study to analysing two organisations within the retail sector, and we focused on the process of return management. Return management encompasses all activities related to returns such as: avoidance, gatekeeping, reverse logistics, and disposal (Rogers et al., 2002). The reason for our focus is that return management involves several decision-making activities related to return requests, reverse logistics, customer relationships, and the optimising of transports. Moreover, return management is expensive for organisations to handle and the reduction of co2-emission due to increased transports is of global importance. The following section briefly presents the HCAI approach. After that, we provide a literature review concerning previous design knowledge regarding the development of DSS involving the combination of HI and AI. Next, we will introduce the research method, followed by a presentation of the results. Finally, we provide a discussion and our conclusions are drawn.

2 Human-Centred AI

HCAI is an emerging scientific field that is based on a socio-technical perspective. It is defined by systems continuously improving from human input and machine learning (Cognizant, 2021). Moreover, results and experiences are based on an effective involvement of both humans and AI (ibid.). Shneiderman (2020) state that one purpose of HCAI is to amplify (i.e., augment and enhance) human performance instead of automating it. A point of departure of HCAI is that, human beings are placed in the foreground, and AI should support humans in making decisions. Shneiderman (2021a) formulates the standpoint of humans more conspicuously when he argues that AI should enhance and empower – not replace – humans. Moreover, Ehsan and Riedl (2020) emphasise the socially situated nature of AI systems and the need for a socio-technical perspective when developing digital support. In addition, Cognizant (2021) outlines the three benefits of HCAI:

- *Informed decision-making*: The combination of the precision of machine learning with human input and values. HCAI enables organisations to make more informed decisions.
- *Reliability and scalability*: The purpose of AI is to help humans and organisations, but without human input and understanding, this help will be restricted. The HCAI approach responds to some criticism of AI by adding emotional input and cognitive knowledge, which means that data becomes more reliable and can be scaled to serve larger needs.
- *More successful software and product-building*: HCAI applies principles of behavioural science to technology. This means that software developers are able to integrate user behaviour into advanced technological solutions.

3 Literature Review

In order to identify previous design knowledge concerning the development of DSS for decision making and involving the combination of HI and AI, we have conducted a literature review. We used the Scopus database and formulated the search string as follows: "human-centered AI" and "design principles" or "design knowledge". The search string returned 18 articles. Out of these we found that:

- Four articles presented design principles related to the co-existence of humans and AI.
- Three articles presented a one-sided focus on AI.
- Eleven articles did not report findings that corresponded to the purpose of our study.

Due to this relatively low number of relevant articles, we also applied backwards reference searching (sometimes called snowball sampling) by reviewing relevant articles cited in the articles identified in the Scopus database. Noy et al. (2008) argue that snowball sampling is the most widely employed sampling method within qualitative research in various disciplines across the social sciences. The backward reference searching resulted in four articles that were relevant to our study. Below, we describe articles that are relevant to our research question.

Subramonyam et al. (2019) have developed several design principles for prototyping AI experiences to support the tasks of end-users. The design principles are oriented towards user experiences (UX) and read: prototyping tools should allow designers to invoke machine learning models by specifying input data directly, prototyping tools should allow designers to incorporate AI outputs into interface design, prototyping tools should allow designers to shape model APIs according to end-user needs, prototyping tools should allow designers to evaluate design choices across diverse users and usage contexts, and prototyping tools should allow flexibility for designers to incorporate model-related data rapidly and iteratively.

Sun et a. (2021) state that the transparency (e.g., explainability and trust) of AI is of highest importance. The authors have conducted a literature study in order to identify articles concerned with interaction issues between AI and humans, such as the transparency of AI agents. The authors propose six levels of transparency for designing transparent AI agents. The levels range from no automation to full automation.

Trakunphutthirak and Lee (2021) state that, in many tasks, machine learning systems have accomplished human performance. The authors present three areas where machine learning is superior to that of humans: the increasing volume of enormous data, improvements in hardware performance and improvements in optimisation algorithms. However, they recognise that some state-of-the-art machine learning techniques, such as deep learning, have been criticised for lacking robust answers and for not being trustworthy.

Jensen (2021) studied the intertwined relationship between trust and anthropomorphism and its implications for designing human-centred AI systems. He argues that measuring trustworthiness and anthropomorphism throughout the design process, alongside performance indicators or metrics of user understanding, can ensure that these ingrained perceptual processes are being sufficiently considered in AI system design.

Cirqueria et al. (2021) have developed design principles for user-centric explainable AI in fraud detection and in relation to DSS. They argue that fraud experts lack trust in AI predictions and have, from a user-centric perspective, suggested five design principles for developing decision support to be used for fraud detection. Unfortunately, the formulation of the design principles is quite lengthy, which is why we did not have room to include them in this paper.

Shneiderman (2020) states that the application of AI leads to high expectations within several sectors. Moreover, he argues that there are many examples of out-of-control robots and biased decision-making. One purpose of the article is to bridge the gap between the ethical principles of HCAI and practical steps for effective governance. He suggests 15 recommendations at three levels: team, organisation, and industry. The purpose of the recommendations is to increase the reliability, safety, and trustworthiness of HCAI systems.

Shneiderman (2021) criticises existing principles concerning the development of AI for being too general and exemplifies by using the following principles: "mitigate social biases" and "AI systems must be transparent and explainable". Furthermore, he argues that specific guidelines are needed for software engineers to perform competently. The article presents three HCAI guidelines: building reliable and transparent systems based on sound software engineering team practices, pursuing safety culture through effective business management strategies, and increasing trust through certification and independent oversight within each industry.

Xu and Dainoff (2021) have identified new challenges related to human interaction with AI systems. The challenges are related to human-controlled AI, human-driven decision-making, explainable AI, usable AI, and ethical & responsible AI. Their conclusion is that current methods are limited in their support of the development of HCAI systems.

In summary, we have found that all the reviewed articles provided insights of great value to our study. However, some remarks need to be made:

- There are only a few articles that have developed design principles addressing the combination of humans and AI with regard to decision making.
- We have not found any article that presents design principles for DSS involving HI and AI in the area of return management.
- Only a few studies offer empirical evidence.

Based on the literature review, we can conclude that there is a need for design knowledge concerning the development of DSS, based on a combination of HI and AI.

4 Research Method

In order to fulfil the purpose of our study, we used Action-Design Research (ADR) (Sein et al., 2011). ADR is a widely used method within the design science research (DSR) paradigm. Sein et al. (2011, p.40) state that "ADR is a research method for generating prescriptive design knowledge through building and evaluating ensemble IT artifacts in an organizational setting". Furthermore, Sein et al. (2011, p.53) state that the ADR method "…provides methodological guidance for researchers who study the design of ensemble artifacts". The ADR method consists of four stages: Problem Formulation, Building, Intervention and Evaluation, Reflection and Learning, and Formalising of Learning (see Figure 1). For each stage there are a number of principles that encapsulate underlying beliefs and values. Due to lack of space, we refer to Sein et al. (2011) for an exhaustive description of the ADR method. In the following sections, we describe our actions related to each phase and associated principles.

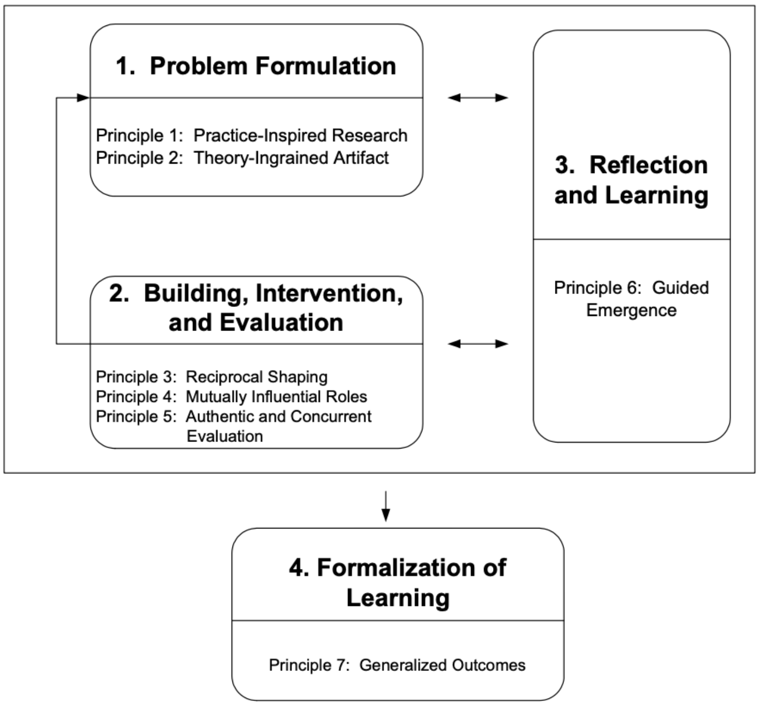

1. Problem Formulation
Principle 1: Practice-Inspired Research
Principle 2: Theory-Ingrained Artifact
2. Building, Intervention, and Evaluation
Principle 3: Reciprocal Shaping
Principle 4: Mutually Influential Roles
Principle 5: Authentic and Concurrent Evaluation
3. Reflection and Learning
Principle 6: Guided Emergence
4. Formalization of Learning
Principle 7: Generalized Outcomes

Figure 1. The stages of the ADR method (Sein et al., 2011).

Problem Formulation. The Problem Formulation stage aims to identify and conceptualise research opportunities based on a problem experienced in practice and theory. First, we organised workshops and conducted interviews in order to identify and formulate the problem experienced in the two participating organisations. Second, we consulted theory in order to learn what is known about the problem and prior advances in technology. These two actions identified a knowledge gap regarding the design of DSS based on a combination of HI and AI (see sections 1-3). The two actions correspond to the two ADR principles "Practice-Inspired Research" and "Theory-Ingrained Artefact".

Building, Intervention and Evaluation (BIE). The stage aims to realise the design of an artefact (i.e., the DSS) and articulate the design principles. In order to build and evaluate the DSS, we followed the ADR principle "Authentic and Concurrent Evaluation". We applied the organisation-dominant BIE form to conduct a naturalistic evaluation inspired by the evaluation framework suggested by Venable et al. (2016). We applied the evaluation criteria utility and fit-for-context. Empirical evidence was collected by organising evaluation sessions and intervening in the two participating organisations. The evaluation sessions consisted of three steps: a) introduction of the DSS to the retail organisation, b) evaluation of the DSS in a real empirical situation involving practitioners with different roles related to return management, and c) collection of empirical evidence (notes were taken during the evaluation session, interviews with the practitioners). We also followed the ADR principle of "Mutual Influential Roles" during the BIE stage, meaning that researchers and practitioners shared theoretical and business knowledge. The results of the evaluation sessions were analysed and specified as new requirements for refining the DSS. Each evaluation session involved two researchers and two-three practitioners. In total, we conducted two iterations, and each iteration included two sessions with both organisations.

During the BIE stage, we also utilised the principle of "Reciprocal Shaping" which addresses the relationship between the development of the DSS and design principles influenced by the organisational context. The articulation of the design principles followed the iterative development of the DSS. This meant that the evaluation of the DSS also provided valuable knowledge to refine the design principles. Moreover, each evaluation session resulted in a need to consult theory. Consequently, the development of the design principles and DSS was based on theoretical insights and empirical evaluations (see Figure 2). We utilised the mutual dependency between the evolving DSS and emerging design principles in the following way:

A) The development of the DSS was guided by the design principles that emerged during the BIE cycles. That is, the advances of the design principles were used to shape the DSS.

B) The development of the design principles was guided by empirical feedback from the use of the DSS. That is, the DSS provided a platform for the evaluation of the design principles.

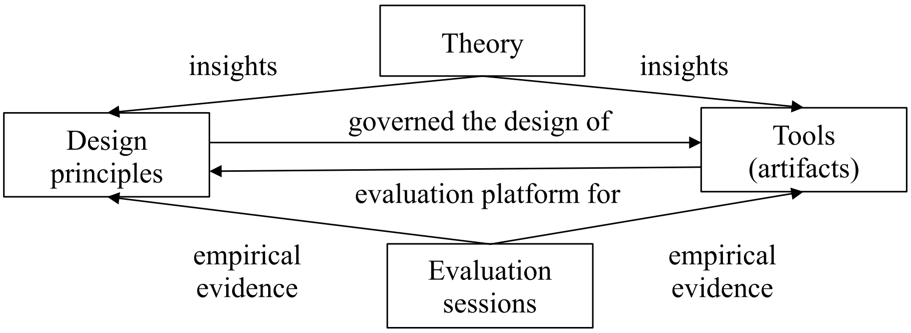


Figure 2. The mutual dependency of the development of design principles and DSS.

Reflection and Learning. This stage aims to move conceptually from building a solution for a particular instance to applying that learning to a broader class of problems. The developed DSS can be seen as an instance of the class "decision-support systems for return management". In order to broaden the learning, we followed the ADR principle "Guided Emergence", emphasising that the artifact will reflect not only the initial requirements but also other instances of this class. To make the learning applicable for a broader class of problems, we analysed the ongoing shaping through organizational use, perspectives, and participants. We also utilised the fact that two organisations participated in the project and searched for similarities and differences between the organisations' contexts. The output from this stage was a generalised solution (the DSS) and generalised design principles.

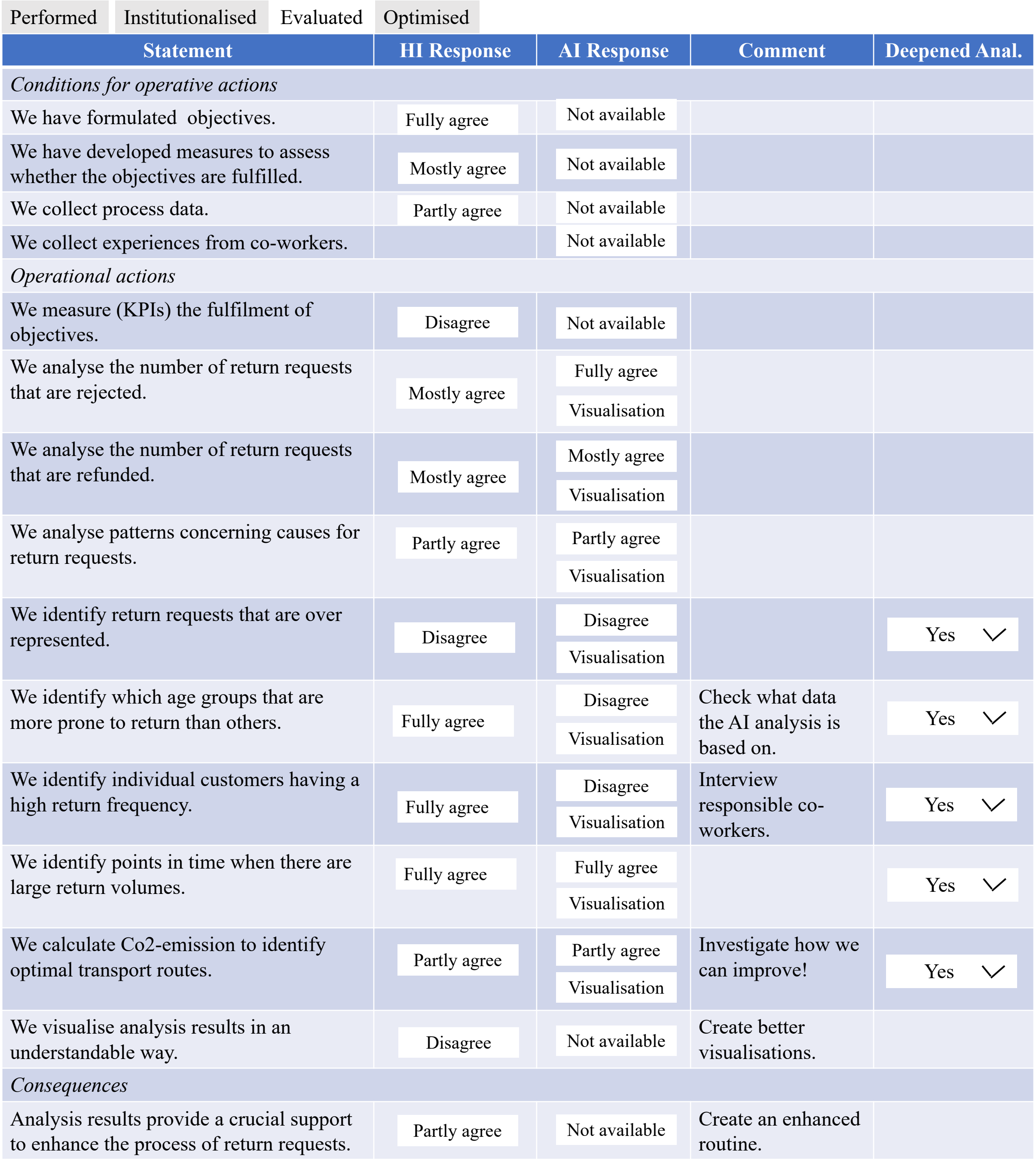
Formalisation of Learning. This stage aims to develop general solution concepts based on the situated learning from an ADR project. To support this stage, we followed the ADR principle "Generalised Outcomes". In order to formalise the design principles in a structured way, we followed the suggestion by Van den Akker (1999): "If you want to design intervention X [for the purpose/function Y in context Z], then you are best advised to give that intervention the characteristics A, B, and C [substantive emphasis], and to do that via procedures K, L, and M [procedural emphasis], because of arguments P, Q, and R.". The reason for choosing this suggestion is that it provides support for richer descriptions of design principles compared to other suggestions that only contain support for formulating the causal relationship "goal → action" (e.g., Walls et al., 1992; Goldkuhl 2004; Van Aken, 2004). The results involving the general solution concepts were presented in research papers (such as this paper), a report required by the funder of the research project, which was also made publicly available to organisations.

5 Results

Section begins with a presentation of the DSS (see section 5.1) which is followed by a description of the design principles (see section 5.2). The reason for presenting the DSS first is to provide a background and context for the suggested design principles.

5.1 Brief Presentation of the DSS

The main idea of the developed DSS is that human beings and AI respond to statements in order to make decisions about enhancements of the return management process. The statements are derived from theory concerning return management and input from the organisations during the evaluation sessions (see figure 3). The participants (the humans) make a consensus-based response by selecting one of the options: fully agree, mostly agree, partly agree, disagree and N/A.

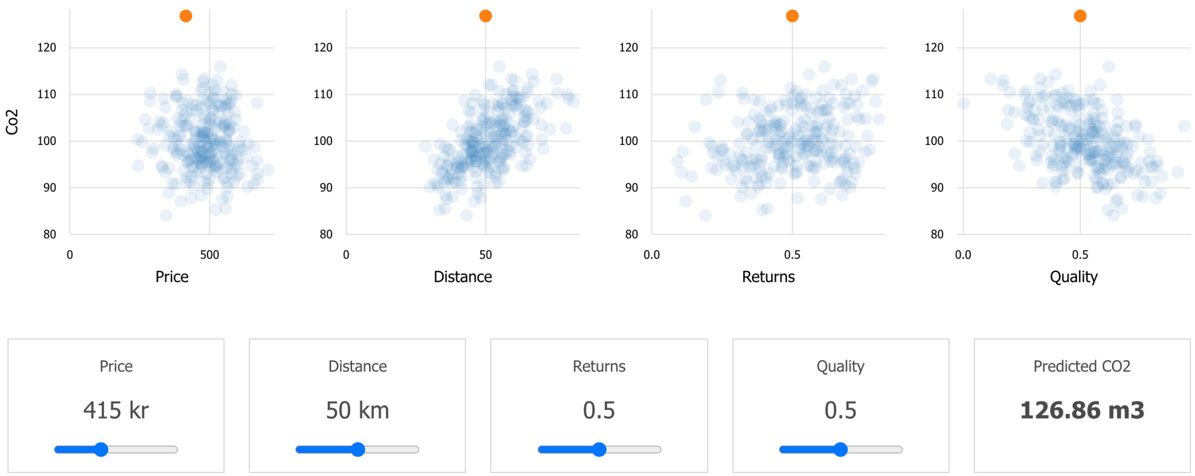


Performed | Institutionalised | Evaluated | Optimised

Statement	HI Response	AI Response	Comment	Deepened Anal.
Conditions for operative actions				
We have formulated objectives.	Fully agree	Not available		
We have developed measures to assess whether the objectives are fulfilled.	Mostly agree	Not available		
We collect process data.	Partly agree	Not available		
We collect experiences from co-workers.		Not available		
Operational actions				
We measure (KPIs) the fulfilment of objectives.	Disagree	Not available		
We analyse the number of return requests that are rejected.	Mostly agree	Fully agree Visualisation		
We analyse the number of return requests that are refunded.	Mostly agree	Mostly agree Visualisation		
We analyse patterns concerning causes for return requests.	Partly agree	Partly agree Visualisation		
We identify return requests that are over represented.	Disagree	Disagree Visualisation		Yes ⌄
We identify which age groups that are more prone to return than others.	Fully agree	Disagree Visualisation	Check what data the AI analysis is based on.	Yes ⌄
We identify individual customers having a high return frequency.	Fully agree	Disagree Visualisation	Interview responsible co-workers.	Yes ⌄
We identify points in time when there are large return volumes.	Fully agree	Fully agree Visualisation		Yes ⌄
We calculate Co2-emission to identify optimal transport routes.	Partly agree	Partly agree Visualisation	Investigate how we can improve!	Yes ⌄
We visualise analysis results in an understandable way.	Disagree	Not available	Create better visualisations.	
Consequences				
Analysis results provide a crucial support to enhance the process of return requests.	Partly agree	Not available	Create an enhanced routine.	

Figure 3. Analysis of statements concerning return management.

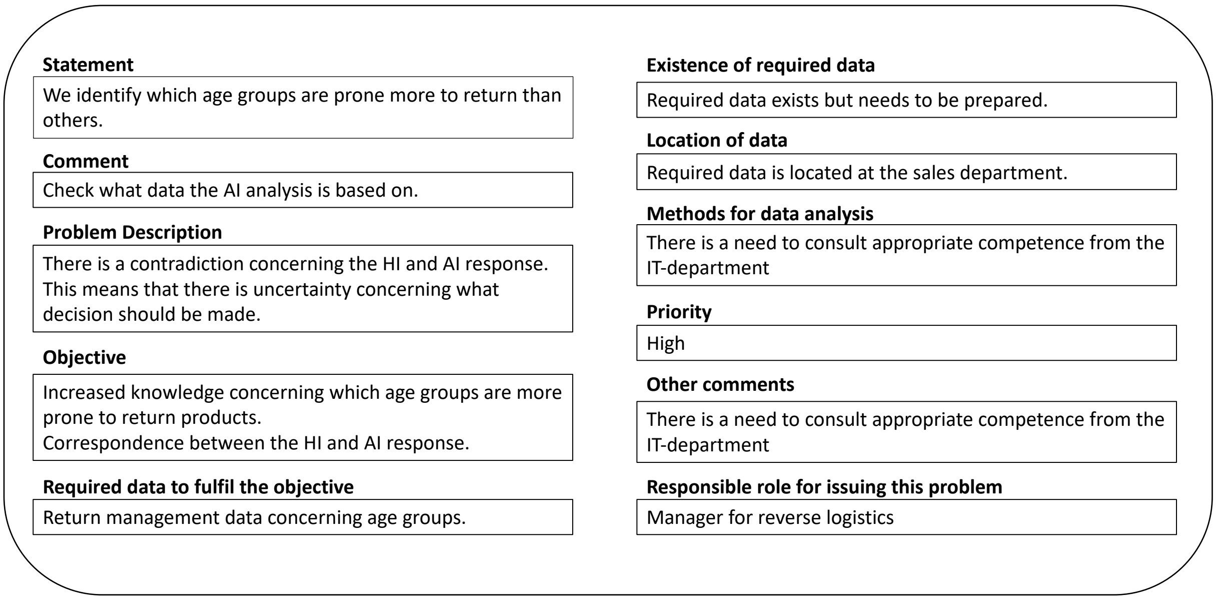
The participants can also add a comment regarding the selected option. After the participants have responded to a statement, they can check the AI response and get a transparent breakdown of the AI response supported by a visualisation by clicking on the button “Show visualisation”. The visualised example refers to the statement “We calculate Co2-emission to identify optimal transport routes” (see figure 4). The visualisation is interactive and the participants can adjust price, distance, returns and quality in order to see how the C02-emission is predicted. We have only implemented AI responses for a few statements since a) we have developed a prototype, and b) AI responses are not meaningful for all statements. For these statements, we have included a text saying “Not available”.



Price	Distance	Returns	Quality	Predicted CO2
415 kr	50 km	0.5	0.5	**126.86 m3**

Figure 4. *Visualised breakdown of the AI response.*

Based on the combined HI and AI responses, the participants can decide whether a deepened analysis is necessary in order to improve specific tasks. The deepened analysis mainly aims to get a broader understanding of the problem, formulate a goal, and decide what data is required to fulfil the goal. The overall idea is to create an action plan for improving the process of return management (see figure 5).



Statement

We identify which age groups are prone more to return than others.

Comment

Check what data the AI analysis is based on.

Problem Description

There is a contradiction concerning the HI and AI response. This means that there is uncertainty concerning what decision should be made.

Objective

Increased knowledge concerning which age groups are more prone to return products.
Correspondence between the HI and AI response.

Required data to fulfil the objective

Return management data concerning age groups.

Existence of required data

Required data exists but needs to be prepared.

Location of data

Required data is located at the sales department.

Methods for data analysis

There is a need to consult appropriate competence from the IT-department

Priority

High

Other comments

There is a need to consult appropriate competence from the IT-department

Responsible role for issuing this problem

Manager for reverse logistics

Figure 5. *Deepened analysis.*

Furthermore, we were inspired by CMMI for Services, developed by The Software Engineering Institute (2010), and arranged the statements according to four capability levels: performed, institutionalised, evaluated and optimised. The “performed level” is the lowest level and the “optimised level” is the highest. The reason for using capability levels was that they provided a) an excellent way to structure the identified statements and b) a self-evaluation approach which meant that the organisations could measure their capability level over time.

5.2 Design Principles

We present three design principles supporting the development of DSS for decision-making, based on a combination of HI and AI, within the area of return management. As mentioned in section 4, the formulation of the design principles follows the suggestions made by Van den Akker (1999).

Design principle 1 – Design for amplified decision-making

If you want to design a digital DSS for the purpose of HCAI in the context of return management, then you are best advised to amplify the DSS with procedural support enabling the exploitation of both HI and AI, because the combination of the strengths of HI and AI provide enhanced decision bases.

Exploiting HI and AI means combining the knowledge, experience and cognitive processes of humans with AI's capabilities of conducting descriptive, predictive or prescriptive analytics. This meant that we designed the DSS to support the possibility for both humans and AI to respond to crucial statements concerning return management. The empirical analysis identified that HI and AI responses to statements could correspond to or contradict one another. The combination of possible HI and AI responses is complex and requires an example. Imagine that an organisation would like to know whether they can identify age groups that are more prone to return products than others. The response based on human knowledge and experiences can be positive (fully agree, mostly agree) or negative (partly agree, disagree). The AI response can also be positive or negative. In the DSS, there is room for providing more detailed responses. However, such details are not relevant when exemplifying the complexity of the combined responses.

The possible responses from humans and from AI give us four possible combinations, which are all essential from the perspective of designing a DSS for amplified decision-making:

A) A positive response from both HI and AI means that the responses correspond to one another and therefore the probability of making the right decision is high.

B) A negative response from both HI and the AI also means that the responses correspond to one another and therefore the probability of making the right decision is high. However, in this case, the character of the decision made is different from the decision made above.

C) A positive response from HI and a negative response from AI means that the responses contradict one another. In this case, the decision basis is inconclusive. Therefore, it is essential to dig deeper in order to understand why there is a mismatch between the HI and AI responses.

D) A negative response from HI and a positive response from AI also means that the responses contradict each other and that there is a need to dig deeper.

The empirical evidence emphasised the importance of designing for amplified decision-making consisting of a combination of both HI and AI. A quote from one of the organisations reads: "I like that HI brings contextual information and experience while AI brings facts; the two types of intelligence complement each other, and when combined, they form superior decision bases compared to each intelligence on its own".

Design Principle 2 – Design for unbiased decision-making

If you want to design a digital DSS for the purpose of HCAI in the context of return management, then you are best advised to give the DSS procedural support enabling the involvement of several roles and consensus-based decision-making, because multiple perspectives are likely to reduce unbiases, uncertainty and vagueness of human judgement.

The empirical analysis identified that the return management involved multiple functional roles across organisational units such as managers for reverse logistics, customer service, warehouses, and sales. This meant that we designed the DSS to encourage participation from multiple perspectives in order to reduce biases against responses to the statements. During the evaluation sessions, it was not unusual for the participants to disagree initially about the formulation of responses to statements. However, we found that the discussions involved a lot of sharing of knowledge and experiences among the participants and that the responses to the statements were made in consensus. A consensus-based process means applying a cooperative process involving all the perspectives presented by different roles leading to an agreement supported by all the roles (DeGroot, 1974). It is a cooperative process in which all group members develop and agree to support a decision in the best interest of the whole (Dressler, 2006). One consequence of the consensus-based process was that the possible biases in responses were reduced. As mentioned above, we also included AI responses to statements. Since we developed a DSS prototype, the AI responses were based on simulated data. However, the participants agreed that the simulated AI response and the data visualisation could reduce the biases in decision-making.

Furthermore, in order to avoid biases among the participants when responding to statements, we found it essential that the human response had been formulated before the humans checked the AI response. Consequently, the AI response was hidden during the discussions among the participants. In order to check how the AI responded to a statement, a click on a button was required. The combination of human and AI responses meant that the AI response was regarded as being part of the consensus process. We found that the design of the DSS reduced the uncertainty and vagueness of human judgement. A quote from one of the organisations reads: "Since multiple perspectives are taken into account, this type of DSS reduces the risk of making decisions based on anecdotes, and therefore I trust these consensus-based decisions more."

Design Principle 3 – Design for human and AI learning

If you want to design a digital DSS for the purpose of HCAI in the context of return management, then you are best advised to give the DSS procedural support enabling both mutual human and machine learning, because knowledge sharing between humans and machines will be likely to improve the organisations' overall organisational capability and engender competitive advantages.

As mentioned above, we discovered that consensus-based decision-making was important in order to formulate responses to statements that would be satisfying to everyone. We also found that sharing knowledge between the participants supported learning about why certain decisions were made. Above, we have discussed the complexity of different combinations of human and AI responses (see design principle 1). The possibility of comparing responses to statements from humans and AI constituted a learning opportunity for both humans and the machine. In the empirical research project, we identified that the participants critically analysed AI responses in order to learn more about specific statements.

We also realised that human knowledge and experiences could be an essential source of information for machine learning. Machine learning can broadly be defined as computational methods using information to improve or make accurate predictions (Mohri et al., 2018). However, machine learning is dependent on the quality of the algorithms and business data (Redman, 2018). In the empirical research project, we observed that when there were differences in human and AI responses, the humans questioned the data quality in terms of: the kind of data the AI response had been built on and whether the data available was sufficient to make reliable predictions. In addition, there was a request for transparency and explanations in order for particular AI responses to be understood. The questioning meant that the humans could identify deficiencies in existing data and suggest complementing data sets that could be included in the AI analyses.

These observations meant that we designed the DSS to include support for both human learning and machine learning. The DSS supported transparency through the provision of decompositions of AI responses. The DSS also provided possibilities to suggest additional data sets to be included in future AI analyses. To summarise, individual participants learned from other participants and from AI responses, and the machine learnt from human input. A quote from one of the organisations reads: "The statements in the DSS helped us to focus on aspects that we did not focus on or discussed before. The DSS also helped us to share knowledge between each other. When interpreting and analysing all the new information with the knowledge provided by the AI response, we could generate even more new knowledge and store it in the database for future use. It's a winning streak".

6 Discussion

This section aims to discuss implications from the implemented design principles. The first design principle, design for amplified decision-making, implies that HI that should be amplified with AI, not the other way around. The formulation follows the intention of HCAI, which states that AI should enhance and empower – not replace - humans Shneiderman (2021a), (see section 2). Moreover, the formulation of the design principle can also be seen as a reaction against other AI-related concepts such as explainable AI, responsible AI, accountable AI and trustworthy AI. All these concepts emanate from a technical AI perspective. One implication of positioning HI in the foreground is that the risk of overlooking human capabilities (e.g., creativity, contextual understanding and common-sense reasoning), will be reduced in decision-making situations. Van der Alst (2021) discusses the relationship between HI and AI in terms of two opposite views: "machine intelligence in the loop of human intelligence" and "human intelligence in the loop of machine intelligence". He points out several advantages and disadvantages with both views and concludes that we need a mixture of human and machine intelligence to get the best results. We agree with this statement because the context of return management involves complexity and diversity. Our experiences from the empirical evaluations are that complexity and diversity is best managed by a combination of HI and AI.

The main message concerning the second design principle, design for unbiassed decision-making, is that the design of the DSS should allow for the participation of multiple functional roles across organisational units. The implication of this design principle is that possible biases will be reduced when knowledge and experiences gained from different units of an organisation are discussed. This implication means that the sharing of knowledge supports a move from knowledge existing at unit levels to organisational learning.

The third design principle, design for human and AI learning, focuses on knowledge sharing between humans and AI. The implication of this design principles is that both humans and AI, in a reciprocal interaction, can learn from each other. The design of the DSS encouraged the humans to critically analyse responses from AI in order to both improve their learning and to improve the AI analysis. Grønsund and Aanestad (2020) have studied the configuration of humans and algorithms in order to find out implications for work and organisation. They are discussing learning in terms of reciprocal human–machine augmentation. One conclusion from their study is that the interplay between humans and algorithms augments one and another. This finding is similar to what we call design for human and AI learning.

As mentioned in section 1, there is a knowledge gap regarding DSS development based on a combination of HI and AI. We claim that the developed DSS and design principles can fill this gap. Inspired by Gregor and Hevner (2013), we view the developed DSS and design principles as interrelated. Gregor and Hevner (2013) present three levels of research contribution types in DSR. Level 1 is called "situated implementation of artifact" and the artefact (the DSS) can be regarded as a carrier of design knowledge. Level 2 consists of nascent design theory, which corresponds to the suggested design principles. Level 3 corresponds to a well-developed design theory, which we have not presented. The levels represent different abstraction levels and are strongly interrelated. Because of the interrelated character of the DSS and design principles, we are viewing all the implications described above as being of both practical and theoretical concern.

7 Conclusion

The research question that guided our research project was formulated as follows: How can decision-making systems for return management be designed through the combination of HI and AI? In order to respond to the research question, we have developed a DSS and design knowledge expressed as three design principles. The DSS constitutes our contribution to practice. Based on theoretical insights and empirical evidence, we can conclude that the combination of HI and AI supported decision-making within return management.

The scientific contribution consists of three design principles: design for amplified decision-making, design for unbiased decision-making and design for human and AI learning. We consider the design principles as a contribution to the emerging field of HCAI.

In section 2, we described HCAI and presented three benefits: informed decision-making, reliability and scalability, and also more successful software and product-building. We regard our design principles as a prescriptive guidance for achieving these benefits. The design principle “design for amplified decision-making” supports “informed decision-making” since it enhances human input and values through the precision of AI. The design principle “design for unbiased decision-making” supports “reliability and scalability”. This is since it reduces human uncertainties and vagueness, which means that input will be more reliable and can be scaled to serve larger needs. We view all three design principles as a support for “more successful software and product-building” concerning HCAI as they prescribe some essential features for integrating human knowledge and experience with AI.

Finally, we can conclude that our findings support the necessity of including the cognitive capabilities of human beings into AI-based decision-support systems to achieve a hybrid intelligence. Our findings show that the combination of HI and AI facilitates decision-making when confronted by complex problems, thereby gaining results that are superior to what can be achieved individually. Our conclusions are based on empirical findings from two organisations in the retail sector. However, we cannot foresee any obstacles concerning the application of the design principles in other retail organisations or other sectors that share similar contextual characteristics. As future research, we suggest a broader empirical study that could further advance design knowledge concerning HCAI. We recommend that future research explicitly focus on the usefulness and applicability of the design principles.

References

Beach, L. R. (1993). Broadening the definition of decision making: The role of prechoice screening of options. *Psychological science*, *4*(4), 215-220.

Borst, C. 2016. “Shared Mental Models in Human-Machine Systems” IFAC-PapersOnLine 49-19 (2016) 195–200

Capterra. (2021). Artificial Intelligence Software. https://www.capterra.com/artificial-intelligence-software/. Retrieved Nov, 2021.

Chandra, L., Seidel, S., and Gregor, S. 2015. “Prescriptive knowledge in IS research: Conceptualizing design principles in terms of materiality, action, and boundary conditions”, in *System Sciences (HICSS), 2015 48th Hawaii International Conference on* (pp 4039-4048). IEEE.

Cirqueira, D., Helfert, M., and Bezbradica, M. (2021, July). Towards Design Principles for User-Centric Explainable AI in Fraud Detection. In *International Conference on Human-Computer Interaction* (pp. 21-40). Springer, Cham.

Cognizant. (2021). Human-Centered Artificial Intelligence. https://www.cognizant.com/us/en/glossary/human-centered-ai. Retrieved Oct, 2021.

Cronholm, S., Göbel, H., and Rittgen, P. (2017). Challenges Concerning Data-Driven Innovation. In *The 28th Australasian Conference on Information Systems, Hobart Australia, December 4-6, 2017..*

Cui, H., Rajagopalan, S., and Ward, A. R. (2020). Predicting product return volume using machine learning methods. *European Journal of Operational Research*, *281*(3), 612-627.

DeGroot, M. H. (1974). “Reaching a consensus”. *Journal of the American Statistical Association*, 69(345), 118-121.

Dellermann, D., Ebel, P., Söllner, M., and Leimeister, J. M. (2019b). Hybrid intelligence. *Business and Information Systems Engineering*, *61*, 637-643.

Demartini, G. 2015. “Hybrid human–machine information systems: Challenges and opportunities”, *Computer Networks* 90 (2015) 5–13

Dressler, L. (2006). *Consensus Through Conversations: How to Achieve High-Commitment Decisions*. Berrett-Koehler Publishers.

Ehsan, U., and Riedl, M. O. (2020, July). Human-centered explainable ai: Towards a reflective sociotechnical approach. In *International Conference on Human-Computer Interaction* (pp. 449-466). Springer, Cham.

Evans, J. R., and Lindner, C. H. (2012). Business analytics: the next frontier for decision sciences. *Decision Line*, *43*(2), 4-6

Göbel, H. and Cronholm, S., 2016. Nascent Design Principles Enabling Digital Service Platforms. In: Proceedings of 11th Int. Conference, DESRIST 2016, Newfoundland and Labrador, Canada.

Goldkuhl, G. 2004. “Design Theories in Information Systems – A Need for Multi-Grounding”, *Journal of Information Technology Theory and Application (JITTA)*, (6:2), pp 59-72.

Göranzon, B. (2009). In Swedish: *Det praktiska intellektet - Datoranvändning och yrkeskunnande*. Stockholm: Santerus.

Gregor, S., Hevner, A.R.: Positioning and presenting design science research for maximum impact. Manag. Inf. Syst. Q. 37, 337–355 (2013).

Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *The Journal of Strategic Information Systems*, *29*(2), 101614.

Jensen, T. (2021, July). Disentangling Trust and Anthropomorphism Toward the Design of Human-Centered AI Systems. In *International Conference on Human-Computer Interaction* (pp. 41-58). Springer, Cham.

Johnson, M., and Vera, A. (2019). No AI is an island: The case for teaming intelligence. *AI Magazine*, *40*, 16-28. https://doi.org/10.1609/aimag.v40i1.2842

Jordan, M. I., and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects, Science, *349*(6245), 255-260.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., ... and Xin, D. (2016). Mllib: Machine learning in apache spark. *The Journal of Machine Learning Research*, *17*(1), 1235-1241.

Mohri, M., Rostamizadeh, A., and Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.

Mumford, E. (2006). The story of socio-technical design: Reflections on its successes, failures and potential. *Information systems journal*, *16*(4), 317-342.

Noy, C. (2008). Sampling knowledge: The hermeneutics of snowball sampling in qualitative research. *International Journal of social research methodology*, *11*(4), 327-344.

OECD (2015), Data-Driven Innovation: Big Data for Growth and Well-Being, OECD Publishing, Paris.

Redman T C. (2018). If Your Data Is Bad, Your Machine Learning Tools Are Useless. Harward Business Review, 2. https://hbr.org/2018/04/if-your-data-is-bad-your-machine-learning-tools-are-useless. Retrieved Oct, 2021

Rogers, D.S., Lambert, D.M., Croxton, K.L. and Garcia-Dastugue, S.J. (2002), “The returns management process”, The International Journal of Logistics Management, Vol. 13 No. 2, pp. 1-18.

Russell S. and Norvig P. (2018). Artificial Intelligence: A Modern Approach. Kuala Lumpur, Malaysia; Pearson Education.

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., and Lindgren, R. (2011). “Action Design Research”. *MIS Quarterly*. Vol 35 No 1, pp. 37-56.

Shneiderman, B. (2020). Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy Human-Centered AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *10*(4), 1-31.

Shneiderman, B. (2021a). Human-Centered AI. *Issues in Science and Technology*, *37*(2), 56-61

Shneiderman, B. (2021b). Human-Centered AI: A New Synthesis. In *IFIP Conference on Human-Computer Interaction* (pp. 3-8). Springer, Cham.

Software Engineering Institute. (2010). CMMI for Services ver 1.3. https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=9665. Retrived Nov, 2021.

Subramonyam, H., Seifert, C., and Adar, E. (2021, April). ProtoAI: Model-Informed Prototyping for AI-Powered Interfaces. In *26th International Conference on Intelligent User Interfaces* (pp. 48-58).

Sun, L., Li, Z., Zhang, Y., Liu, Y., Lou, S., and Zhou, Z. (2021, May). Capturing the Trends, Applications, Issues, and Potential Strategies of Designing Transparent AI Agents. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-8).

Trakunphutthirak, R., and Lee, V. C. (2021). Application of Educational Data Mining Approach for Student Academic Performance Prediction Using Progressive Temporal Data. *Journal of Educational Computing Research*.

Van Aken, J. 2004, “Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-tested and Grounded Technological Rules”, *Journal of Management Studies*, (41:2), pp 219-246.

Van den Akker, J. 1999. "Principles and methods of development research". *Design approaches and tools in education and training*, pp. 1–14. Springer.

Van der Aalst, W. M. (2021). Hybrid Intelligence: to automate or not to automate, that is the question. *International Journal of Information Systems and Project Management*, *9*(2), 5-20.

Venable, J., Pries-Heje, J. and Baskerville, R., (2016). "FEDS: a framework for evaluation in design science research", *European Journal of Information Systems,* Vol 25 No 1, pp. 77- 89.

Walls, J.G., Widmeyer, G.R., and El Sawy, O.A. 1992. "Building an Information Systems Design Theory for Vigilant EIS", Information Systems Research (3:1), pp. 36-59.

Xu, W. (2019). Toward human-centered AI: a perspective from human-computer ineraction. *Interactions*, *26*(4), pp. 42-46. https://dl.acm.org/doi/fullHtml/10.1145/3328485?casa_token=PCClX3PzVVsAAAAA:FMfNERuoARXVMoE634xLhZ0LTr39GU8Oa10ChYJUsgQg0tBB2CZi9uvZQ7aiaC73z-8Onk2jzneK.

Xu, W., and Dainoff, M. J. (2021). Transitioning to human interaction with AI systems: New challenges and opportunities for HCI professionals to enable human-centered AI.

Zheng, N., Liu, Z., Ren, P. et al. (2017). Hybrid-augmented intelligence: collaboration and cognition. *Frontiers of Information Technology and Electronic Engineering, 18,* 153-179. https://doi.org/10.1631/FITEE.170005