Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 239 (2024) 563–574

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist – International Conference on Health and Social Care Information Systems and Technologies 2023

Design Principles for Data-Driven Decision Evaluation

Nada Elgendy[a,*], Tero Päivärinta[a,b], Ahmed Elragal[b], Karoliina Hannula[c], Kaisa Puolitaival[c]

[a]M3S, Faculty of Information Technology and Electrical Engineering, University of Oulu, Finland
[b]Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, Sweden
[c]Antell, Oulu, Finland

Abstract

Human-machine collaboration has potentially led to higher quality and more informed data-driven decisions. However, evaluating these decisions is necessary to measure the benefits, as well as enable experiential learning and posterior rationalization of the results and consequences. Nevertheless, the multiplicity of human-machine collaboration modes, as well as the multi-faceted nature of data-driven decisions complicates evaluation, and evaluation solutions are lacking both in research and in practice. This is further reflected in the complexity of incorporating evaluation in the design of such data-driven decision making systems, since developers are left without theoretically grounded and practically feasible principles to guide implementation. In this paper, we propose a set of five design principles, explicated from theory and practice, for systems implementing data-driven decision evaluation as the output of design science research cycles. The design principles are: 1) multi-faceted evaluation criteria, 2) unified viewpoint, 3) collaborative rationality, 4) processual ex-post evaluation, and 5) adaptive feedback and learning loops. They are further contextualized in the case of AI-enabled menu design at Antell, an innovative pioneer in the restaurant business in Finland, and consequently evaluated by the development managers of the project. Accordingly, the design principles contribute to the knowledge base on metahuman systems and data-driven decision evaluation, by concretizing existing normative concepts into prescriptive knowledge, also guiding future research and generalizing towards a design theory. Furthermore, they provide implementable statements for designing and developing such systems in practice and can be used as a checklist to compare and evaluate existing systems.

Peer-review under responsibility of the scientific committee of the CENTERIS / ProjMAN / HCist 2023

Keywords: data-driven decision making; human-machine collaboration; ex-post evaluation; collaborative rationality; design principles.

* Corresponding author.
E-mail address: nada.sanad@oulu.fi

1877-0509
Peer-review under responsibility of the scientific committee of the CENTERIS / ProjMAN / HCist 2023
10.1016/j.procs.2024.06.208

1. Introduction

Advancements in technology and the augmentation of artificial intelligence with human capabilities in organizations has transformed decision making systems, elements, and processes [1, 2]. Data-driven decision making has evolved to encompass not only the traditional decision maker, decision-making process, and the decision, but to also encapsulate voluminous and various types of data which are processed by sophisticated and intelligent machines collaborating with humans in enhanced decision-making systems [1]. While such human-machine collaboration has potentially led to higher quality and more informed collaborative rationality-based decisions, evaluation is crucial to measure its benefits and influence, as well as enable experiential learning and posterior rationalization of the results and consequences. However, the multiplicity of modes of collaboration and interaction between humans and machines inevitably increases the complexity of decision making, consequentially complicating decision evaluation [3–5].

The implementation of data-driven decisions requires more research and should be approached with caution and deliberation. New organizational functions are required for metahuman systems (where humans and machines learn jointly) to *delegate* human-machine decisions and discover what happens if agency is shifted in the system, *monitor* how data-driven decisions are taken and how humans interact with machines, *cultivate* criteria to evaluate decisions and learning in such metahuman systems to serve human defined and monitored processes, and *reflect* through double-loop learning by creating processes to help evaluate such systems [6], while preserving the human-centric nature of decision-making and determining the potential consequences of the decisions [7]. Thus, holistic perspectives for data-driven decisions and the systems implementing them are vital, since limited vision and imbalanced collaboration and augmentation of human-machine capabilities can lead to destruction and havoc to organizations [8], with a multitude of further unintended negative consequences [9]. However, no comprehensive or holistic design solutions accommodating for the multifaceted nature of collaborative data-driven decisions are found in the literature [4, 5].

Accordingly, our research question is: *"What principles guide designing for the evaluation of data-driven decisions?"*. In prior research, the design objectives for a data-driven decision evaluation solution were determined from theory and a practical case [4]. Building on these design objectives and the theoretical underpinnings elucidated in [5] we developed a data-driven decision evaluation model as a design science artifact and evaluated it in multiple iterations in the practical case of AI-enabled menu design at Antell, a restaurant business in Finland. Finally, based on the design science knowledge gained in the previous stages, in this paper we extrapolate a set of theoretically grounded and practically feasible design principles (DPs) and evaluate them within the case of Antell. Developers and organizations can utilize these DPs to design comprehensive and valuable data-driven decision making and evaluation systems. Furthermore, the DPs can be used as a checklist with which to compare and evaluate existing data-driven decision systems. Finally, the DPs add to the theoretical knowledge base as design science research (DSR) contributions and can be used to guide future research on data-driven decision making and evaluation.

The paper is divided as follows. Section 2 explains our proposed DPs, which are contextualized in the case of Antell in Section 3. The DPs are evaluated in Antell in Section 4. Finally, we discuss the feedback received on our DPs as well as their practical and theoretical implications in Section 5, thus concluding the paper.

2. Design Principles as Prescriptive Knowledge

What distinguishes design science knowledge from other types is its provision of prescriptive knowledge in the form of DPs. DPs are prescriptive "know how" statements that can be derived from abstract artifacts, such as models, and indicate what and how to build an artifact to achieve a predefined goal [10, 11]. They are level 2 contributions of nascent design theories in DSR [12]. Thus DPs are important for capturing and communicating essential design knowledge, generalizing prescriptive knowledge, and as steps for developing comprehensive bodies of knowledge or design theories [10]. Accordingly, following Gregor et al.'s [11] anatomy of a DP, we were able to derive five DPs from previous work, elaborated in Tables 1-5.

Table 1. Design principle 1.

Design Principle	Multi-faceted evaluation criteria
Aim, implementer, and user	For developers (implementers) to design comprehensive and valuable data-driven evaluation solutions (aim) for use by human and machine decision makers (users)
Context	In systems implementing data-driven decision evaluation
Mechanism	Incorporate multi-faceted human and machine evaluation criteria and metrics across contextual levels. These criteria should be capable of differentiating between the data-driven decision elements, such as the evaluation of the machine, the decision outcome, the data, etc., which may potentially be conflicting and otherwise lead to confusion. Furthermore, not only machine metrics should be considered, but also metrics defined by the human decision makers, the organization, and the environment
Rationale	Because humans evaluate decisions differently than machines and utilize different metrics and methods. Since data-driven decision making elements include the decision maker, decision making-process, decision, data, and machine, evaluation should not be myopic or limited to a single element. Thus, a set of conformance criteria (ranging across various contextual levels, i.e. decision, organization, and environment) should govern the entire decision, which is different/more than each of the human and machine agents' sets of conformance criteria to which their individual choices conform.

The first DP, described in Table 1, focuses on incorporating multi-faceted evaluation criteria in systems implementing data-driven decision evaluation [4]. Data-driven decisions comprise of five main elements, the decision maker, decision-making process, decisions, data, and the machine [1]. Each of these elements is evaluated from various perspectives, and humans perform evaluation differently than machines [5]. This requires utilizing a wider range of conformance criteria and metrics from each of the decision, organizational, and environmental contexts [4].

Table 2. Design principle 2.

Design Principle	Unified viewpoint
Aim, implementer, and user	For developers (implementers) to facilitate informed, standardized, and collaborative decision making (aim) for decision makers (users) as well as enhance and inform the iterative development process (aim)
Context	In systems implementing data-driven decision evaluation
Mechanism	Model and document a unified viewpoint of the decision, its elements, and its evaluation
Rationale	Because each of the humans and machines involved in decision making have different, and potentially conflicting, goals and objectives, values, and views of the decision and its evaluation. A unified viewpoint and holistic understanding of the data-driven decision, its elements, and the evaluation of the decision should be made visible to each of the decision makers involved in the decision making. Accordingly, this would allow a similar understanding of the decision goals and objectives, the data utilized, the responsibilities and relationship between the human and machine, the decision making process, and how the decision is evaluated.

The second DP, in Table 2, suggests that systems implementing data-driven decision evaluation should facilitate the modelling and documentation of a unified viewpoint of the decision, its elements, and its evaluation. Humans and machines have different views of the decision, its context, and evaluation [5], and different human decision makers are likely to frame the same decision differently [13]. Without clarifying the evaluation grounds for the decision, stakeholders and decision makers (human and machine) are likely to evaluate the decision differently, leading to conflicting perspectives and processes which enable resistance and limit the benefits realization of the system. Accordingly, providing a unified viewpoint would allow a more comprehensive and holistic understanding of the entire decision between all decision makers involved, and a standardized decision making and evaluation process.

Table 3. Design principle 3.

Design Principle	Collaborative rationality
Aim, implementer, and user	For developers (implementers) to facilitate more coordinated human-machine collaboration for decision makers (users) leading to better quality decisions (aim)
Context	In systems implementing data-driven decision evaluation
Mechanism	Define collaborative rationality between humans and machines and the mode of collaboration between them in decision-making
Rationale	Because collaborative rationality jointly combines the capabilities of humans and machines, based on the available data, to solve problems and learn together. It is more than an aggregation of individual rationalities, but is rather a complementary entity of its own, characterized by its own features. With such definition, more insights can be provided on how humans and machines can collaborate and its effect on decision making and decision outcomes, thus steering the collaboration to be more coordinated and resulting in better quality decisions.

The third DP, explained in Table 3, proposes that collaborative rationality between humans and machines, and the mode of collaboration between them should be defined. There are multiple modes for human-machine collaboration. For example, either AI decides and implements with humans only providing supervision and oversight, AI decides and the human implements, AI recommends and the human decides, AI generates insights which the human uses in the decision process, or the human generates hypothetical situations and relies on AI to evaluate and assess them [14]. Thus the decisions can be either be made purely by the machine, be sequence-based where the human delegates to the machine or the machine delegates to the human, or be aggregated resulting from human-machine joint work [15].

Thus, human-machine interaction and roles need to be clarified and monitored [6] in order to evaluate the resulting decisions and determine the benefit, impact, and learning that consequently occur from this collaboration [1, 3, 16]. Accordingly, more insights can be provided on the effect of collaboration on decision making, making it more coordinated, and users can be more willing to adopt the system with the clarification of roles.

Table 4. Design principle 4.

Design Principle	Processual ex-post evaluation
Aim, implementer, and user	For developers (implementers) to enable continuous experiential learning, posterior rationalization, and sensemaking for human and machine decision makers (users) from the actual consequences and outcomes of the data-driven decision, as well as capture the resulting changes across time in order to enhance future decisions (aim)
Context	In systems implementing data-driven decision evaluation
Mechanism	Facilitate the performance of processual, ex-post evaluation across different stages in time
Rationale	Because ex-post evaluation is generally triggered by the occurrence of particular events or markers, and can be conducted in a processual manner across time to encapsulate changes in the environment, changing contexts, and aspects regarding the data-driven decision and its outcomes. Such changes should be captured in the evaluation to understand the longitudinal effects and consequences of the decision.

The fourth DP, elaborated in Table 4, explicates the need for processual [17, 18] ex-post evaluation to be facilitated in such systems in order to encapsulate changes in the environment and influence them in the desired direction over time. This aids in understanding the emergent, situational, and holistic features of the decision, or the decision-making process, in its changing context [19], and enables experiential learning, posterior rationalization, and sensemaking from the outcomes and consequences of the decision [5].

Finally, the fifth DP, shown in Table 5, is for adaptive feedback and learning loops. The core value of ex-post evaluation is the feedback which drives the new information and knowledge back into the loop, resulting in different types of learning, such as single-loop, double-loop, and deutero learning [20]. These feedback loops should evolve, through experiential learning, to adapt to the changing context and requirements across time. They may (at least initially) exhibit user initiated adaptability, where the developers and decision makers are in control of the system parameters, or preferably evolve to more automated system adaptivity [21].

Table 5. Design principle 5.

Design Principle	**Adaptive feedback and learning loops**
Aim, implementer, and user	For developers (implementers) to enable single-loop, double-loop, and deutero learning for human and machine decision makers (users) from the feedback provided by evaluating past data-driven decisions in order to enhance future decisions (aim)
Context	In systems implementing data-driven decision evaluation
Mechanism	Enable (potentially automated) adaptive feedback loops for learning from the (discrete or continuous) evaluation of past decisions. The feedback loops should adapt and evolve according to the changing context and requirements across time
Rationale	Because with the inferences from the feedback provided by the ex-post evaluation, not only single-loop learning, but also double-loop learning and deutero learning can occur. Accordingly, we can learn more about collaborative rationality, how to enhance it, and how to improve and augment the learning of humans and machines. This can serve as input to enhance the quality of future decisions.

3. Contextualizing the Design Principles: Antell’s Case of AI-Enabled Menu Design

Antell is a family-owned business, founded in 1880, with approximately 70 lunch and staff restaurants across Finland. The group’s turnover was about 23 million euros in 2022, and it employs around 300 professionals. The company’s head office is located in Oulu, Finland. The restaurant business is a significant sector of tourism and hospitality, which has recently started shifting to innovative automation enabled by AI. However, restaurants are people-driven businesses which previously lacked reliance on organizational IT structures, thus creating obstacles for restaurant managers to manage the move from manual tasks to more automated technological systems. Consequently, technological developments for restaurant operations represent a novel case for data-driven decision making, in which further research is required on the collaboration between humans and machines to create innovative processes, new forms of customer experiences, and networked interactions in restaurant ecosystems [22].

The choice of menu items is an important decision for a restaurant’s operations, where the combination of ingredients is an influential attribute in determining client appeal, and menu item specifications are modified to optimize profitability. Menu listings dictate most of a restaurant’s decisions in purchasing, production, and services. Accordingly, menu analysis supports decision making in modifying existing menu items, or introducing new ones, by evaluating the performance of menu items for increasing the contribution margin of each item [23, 24].

At Antell, each restaurant is given a personalized lunch buffet menu that accounts for the needs of the restaurant's clientele. The decisions on the contents of the menus are influenced by the terms of the restaurant contracts and Antell's objectives. Consequently, the menu is developed in collaboration between the product development team and the restaurant manager to ensure that the menus are in accordance with the wishes of Antell and the restaurant’s clientele.

In early 2021, the development of a novel AI tool, given the name “ATSO”, started with the aim of (to an extent) automating the menu decision making process and creating varied, diverse, and cost-effective menus tailored to each restaurant. It was also designed to reduce the amount of manual work by creating optimal restaurant-specific menus

Table 6. Problem and solution statements and their corresponding design principles.

No.	Problem Statement	Solution Statement	Design Principle
1	Humans evaluate decisions differently than machines and utilize limited evaluation criteria. Evaluation has a myopic view based on the evaluator, human judgement and bias play a large role, and global conformance criteria are not clear to all the stakeholders.	A set of global conformance criteria (ranging across various contextual levels) were clearly defined to govern the entire decision. Such criteria are different/more than each of the human and machine agents' sets of conformance criteria to which their individual choices conform. New metrics were specified, including machine metrics, which were necessary for evaluation.	**Multi-faceted evaluation criteria and metrics**
2	Each of the humans and machines involved in decision making have different, and potentially conflicting, goals and objectives, values, and views for the decision and its evaluation. Each decision maker had their own viewpoint of the decision and evaluation based on their individual knowledge, differing from the viewpoints of other decision makers as well as the limited contextual perception of the machine.	Framing the evaluation of data-driven decisions (as a whole), resulting from collaborative rationality, is different than the framing of individual evaluations of each of the agents involved, and can lead to more holistic insights and learning. The data-driven decision, its elements, and evaluation were framed and modeled as a whole, providing a unified viewpoint to the development team and the restaurant managers and leading to a holistic perspective.	**Unified viewpoint**
3	Humans and machines are each bounded by their own rationalities, and the collaboration between them in decision making is uncoordinated and burdened by multiple challenges and issues. Collaborative rationality and the mode of human-machine collaboration was not previously defined or clearly understood by all decision makers.	Collaborative rationality jointly combines human-machine capabilities, based on the available data, to solve problems and learn together. More than an aggregation of individual rationalities, it is a complementary entity of its own, characterized by its own features. Collaborative rationality was more clearly defined. Decision makers were more aware of their roles and the role of the machine, which led to higher levels of trust, transparency, acceptance, and more coordinated decision-making.	**Collaborative rationality**
4	Ex-post evaluation can be overlooked, and sometimes it is not clear when, why, or how it should be done. Ex-post evaluation was informally conducted, without clearly defined evaluation criteria and metrics. Evaluation was mostly in response to triggers, such as complaints.	The value of ex-post evaluation for enabling experiential learning, posterior rationalization, and sensemaking from the consequences and outcomes of the data-driven decision was realized. Ex-post evaluation could be conducted in a processual manner across time to encapsulate changes in the environment. More formal ex-post evaluation criteria and metrics were defined, as well as data collection methods. The consequences and outcomes of the decision could be better monitored across time and evaluated after their definition. Evaluation could be planned more routinely.	**Processual ex-post evaluation**
5	It is unclear when and how learning occurs from ex-post evaluation. Learning is restricted to certain forms, in particular, single-loop learning where errors and problems were identified and addressed. The collaboration between humans and machines and the mutual learning between both is difficult to determine and enhance.	Various types of learning occur in multiple situations and can increase across time with processual evaluation. With the inferences from the feedback provided by the ex-post evaluation, not only single-loop learning, but also double-loop learning and deutero learning were introduced which can allow intelligent adaptation according to the context. Additionally, more could be understood about collaborative rationality, how to enhance it, and how to improve and augment the learning of humans and machines.	**Adaptive feedback and learning loops**

according to specified features and constraints. ATSO led Antell to becoming a finalist for innovation in PRO Gala, the top hospitality industry recognition award in Finland. The development of ATSO progressed in iterative stages and faced some of the classical challenges typical to designing and implementing data-driven decisions systems. Table 6 summarizes the problem statements faced, the solution statements realized by applying our data-driven decision evaluation model in Antell, and the DP each set of statements correlates to. Each DP is contextualized in Antell's case and elaborated in the subsections below. Due to confidentiality constraints, not all details are disclosed.

3.1. DP1 at Antell: Multi-faceted evaluation criteria and metrics

In the initial stages of ATSO's development and implementation, the evaluation criteria and metrics were less defined, more limited, and differences in human-machine evaluation criteria were not well accommodated for. While ATSO had high accuracy in decreasing costs of dishes, it was prone to error in other contexts which human decision makers could immediately identify based on their knowledge. For example, when asked for a vegetable dish, it initially recommended classical Christmas dishes with vegetables, which are inappropriate at other times of the year, since it does not perceive the entire environment or decision context (such as seasonality of recipes). Overcoming these errors required further definition of the constraints and context of the decision within the AI system. Furthermore, restaurant managers might, implicitly or explicitly, evaluate decisions differently, for example rejecting ATSO's combination of menu items which they consider not to be visually appealing, or color-coordinated, when offered together.

Accordingly, in the consecutive stages of ATSO's implementation, global conformance criteria ranging across the contextual levels were clearly defined to govern the entire decision. New metrics were specified, including machine metrics, which were necessary for evaluation. These multi-faceted criteria evolved to include some strategic variables which extend across the environmental and organizational contexts, such as financial variables, the company's image, and environmental variables (e.g. CO_2 and food loss), as well as variables within the decision context, including the time for recipe preparation, ingredient costs and cost of dish, ingredient availability, menu periods, seasonality of recipes, dish quality, dish variability, balance of menu items, etc. Such criteria are different/more than each of the human and machine agents' sets of conformance criteria to which their individual choices conform.

3.2. DP2 at Antell: Unified viewpoint

In the initial stages of ATSO's development and implementation, each of the decision makers had their own internal models and perceptions of the decision-making process and the menu requirements. Each of their viewpoints differed from those of others based on their experience and knowledge of the data, products, costs, restaurant conditions, context, etc. Consequently, the role of each decision maker was neither clear nor well-defined, and the question of whose perception and decision was correct and should be implemented was left unanswered. The human decision makers were naturally wary of using ATSO and implementing its decisions, partially due to their lack of understanding of how the system works and makes the decisions, their own biases and sense of "knowing better", and lack of a clear and standardized structure and process for data-driven decision making. Furthermore, decision makers were prone to errors and incorrect guesses due to lack of data (e.g. not knowing actual portion sizes and having to guess). Moreover, each decision maker evaluated the decision differently and had different goals and objectives.

In subsequent development iterations of the system, the data-driven decision, its elements, and evaluation were framed and modeled as a whole, providing a unified viewpoint to the development team and the restaurant managers and leading to a more holistic perspective for everyone. Hence, the development team could clearly see the different decision elements and informedly plan any necessary enhancements, which resulted in the restaurant managers' increased understanding of the system and goals, and willingness to utilize ATSO. During contextualization of this DP, we find that the DP is not only useful for decision makers as the users of the system, but also for developers during their iterative implementation and enhancement of the data-driven decision system.

3.3. DP3 at Antell: Collaborative rationality

In the initial stages of ATSO's development and implementation, collaborative rationality was not defined as a separate concept of its own. The mode of collaboration between human decision makers and ATSO was not clear, nor was the role of each. For example, restaurant managers were aversive to trying new foods suggested by ATSO in their restaurant and, based on their personal judgement, believed that customers would not eat such kinds of food. Thus, many decisions were based on emotions and irrational expectations of AI's capabilities. In this stage, decision makers were affected by biases emerging from their lack of trust in the machine and incorrect assumptions in their models of the world. The restaurant managers were also overconfident of their own knowledge and expertise, which biased their choices, and the introduction of ATSO created undesired emotions, such as stress and reluctance to change from their traditional decision processes and professional judgement.

In subsequent stages, collaborative rationality and the mode of collaboration was clearly defined. While ATSO recommends, the restaurant manager decides and is held accountable for the decisions and their results. Although restaurant managers still had a limited understanding of how ATSO works, trust increased in later development iterations when they had more control on the input parameters and could make their own modifications. Furthermore, they started perceiving the value added by ATSO and its decisions, and were thus more willing to utilize it as a tool. Collaborative rationality jointly combines the capabilities of humans and machines, based on the available data, to solve problems and learn together. More than an aggregation of individual rationalities, it is rather a complementary entity of its own, characterized by its own features. Based on March's (1978) models and ideas of rationality, a limited rationality is exhibited where the human cannot anticipate all the alternatives, and the machine cannot perceive the entire context of the decision. Additionally, there is adaptive rationality, where the human can adapt to changes in the restaurant environment, selected rationality where some of the menu choices are dominated by standard operating procedures, and posterior rationality where the decision makers can justify their choices based on posterior consistency of the outcomes. Hence, decision makers were more aware of their roles, and the role of the machine, which led to higher levels of trust, transparency, acceptance, and more coordinated decision-making.

3.4. DP4 at Antell: Processual ex-post evaluation

In the initial stages of ATSO's development and implementation, ex-post evaluation was not comprehensively planned but was prematurely triggered by negative feedback which came from a small group of restaurant managers resistant to change. The data gathered and used in the evaluation included verbal feedback from the restaurant managers in informal discussions, as well as formally collected written feedback. The actual vs. expected cost of the menu was calculated to quantitively evaluate the decision, and customer taste points were determined, based on the actual outcomes observed. Later on, data on the number of customers was incorporated into the evaluation, and the monthly purchase percent was added as a metric.

Furthermore, the impact and consequences of Antell's data-driven decisions needed to be considered across time since they could be immediate, such as with ingredient costs, or delayed, like with customer long-term satisfaction. Additionally, the impact could be direct, such as observing higher levels of customer satisfaction with the menus, or indirect, such as observing B2B customer satisfaction with Antell's model of operation and contract continuation. Consequently, the value of processual ex-post evaluation to encapsulate changes in the environment and monitor and learn from the consequences of the decision across time was realized. This enabled the definition of more formal ex-post evaluation criteria and metrics, as well as data collection methods, and evaluation was planned more routinely.

3.5. DP5 at Antell: Adaptive feedback and learning loops

In the initial stages of ATSO's development and implementation, feedback was limited and was mostly negative. Nevertheless, it led to enhancements in the system to accommodate for the new information learned. With the progressive development of the system, and the more data-driven decisions were made and evaluated, the feedback loops evolved and adapted to provide new information which came from the restaurant managers', product development team's, and customers' feedback, as well as the financial figures. The results or realizations which ensued served as valuable input for the following iterations of the data-driven decision making and evaluation system and various types of learning took place. Thus, both the data-driven decision system and the feedback loops were adapted and enhanced according to the new knowledge and insights gleaned. Additionally, more could be understood about collaborative rationality, how to enhance it, and how to improve and augment the learning of humans and machines.

With the inferences from the feedback, single-loop, double-loop, and deutero learning ensued supporting intelligent adaptation according to the context. Single-loop learning allowed the decision makers to take immediate actions, or contingencies, to address emerging problems as they occurred. This included immediate updating of generated errors in the menu (e.g. in the case of erroneously proposed Christmas dishes), and developmental changes in the system to avoid such errors in the future. Double-loop learning led to updating the governing variables of the decision and organizational norms and processes. For example, it was found that restaurant managers erred in portion sizes since they didn't have data on the food waste resulting from the menu, and that portion sizes should be changed to be based on data. Thus, food waste was added as governing variable to define the effective performance of the menu decisions.

Moreover, through deutero learning, Antell was able to learn more about the collaboration and collaborative rationality ensuing between the human and machine decision makers, and their learning processes. With posterior rationalization and sensemaking with retrospective, decision makers were able to learn from the results and outcomes of the entire decision, as well as what worked, what didn't, and which decision maker was more capable at what. For example, it was realized that with ATSO there was better control over the budget, which was an important governing variable. The variety and variability of menu dishes provided by ATSO was also better than that of humans, since human cognitive capacity and memory does not allow remembering all possible recipes which could be used. Nonetheless, restaurant managers were much more adept at perceiving the entire environment and context of the decision, as well exercising their experience and judgement in making the final decision and understanding the customers of their restaurants. Consequently, Antell could specify the failures in ATSO's learning and act accordingly, as well as pinpoint the failures in human learning in collaborating with ATSO and utilizing its features. Hence, the practices which inhibited human, machine, and collaborative learning were identified, and new strategies and modifications could be planned and implemented, as well as reflected in organizational learning practices. While the system and the feedback and learning loops in Antell are not fully adaptive on their own, further development iterations can results in higher levels of automation.

4. Evaluation of the Design Principles

A post-mortem meeting was held at the Antell headquarters in Oulu, Finland with two development managers (DM). DM 1 has a background in international business and communication, works with processes and projects, and managed the first half of the ATSO AI project. DM 2 has a background in economics, accounting, and information processing science, works with numbers, and managed (and is currently managing) the latter half of project. During the meeting, the DMs were shown the five DPs resulting from our DSR cycles. They were asked to individually evaluate each of the DPs according to Iivari et al.'s [25] design principle evaluation criteria, and write down their evaluation. Subsequently, their evaluations for each of the DPs were compared, discussed in more detail, and the discussions were documented.

Iivari et al. [25] suggest that DPs are evaluated for reusability against five, ordered dimensions. *Accessibility* refers to the way DPs are presented and whether they can be understood and comprehended by the target community, and are individually and collectively intelligible. *Importance* refers to the importance of the real-world problems the DPs help to address and provide solutions to. *Novelty* and insightfulness imply not only scientific, but also practical novelty and that the DPs are seemingly new and innovative. *Actability* and guidance refer to the DPs being realistic and can be acted and carried out in practice, thus providing sufficient guidance. *Effectiveness* refers to the effect of reusing the DPs in the adopting unit (i.e. development process of the system) and its potential relative advantage and usefulness.

During evaluation, if the first criterion is met ("Yes"), the following criterion is checked, etc. However, if it is not met ("No"), the following criteria are deemed irrelevant and the evaluation of the DP ends. In our case, we chose to evaluate how well each DP meets the criteria on a scale of 0-5, where 0 is equivalent to "No", and 1-5 is equivalent to "Yes", with values of 1, 2, 3, 4 and 5 representing "very low", "low", "moderate", "high", "very high" respectively. The reason for selecting a scale as opposed to simple binary evaluation is to allow room for discussion, comparison, improvement, and future research. However, this does not represent any quantitative value or significance, and our evaluation of the DPs was purely qualitative. The design principles were well-received across the different dimensions in relation to Antell's case, with all dimensions for each of the DPs receiving a "moderate", "high", or "very high" opinion, with most receiving "very high". No DPs were evaluated negatively. Accordingly, the DPs can be applied and reused in this context, and potentially in other cases.

During the discussion with the DMs, DP1 was found to be important, novel, effective, and value-adding. It is a crucial part of holistic and comprehensive evaluation and is a facilitator for DPs 4 and 5. To meaningfully perform processual ex-post evaluation and gain beneficial feedback for enabling learning, evaluation should be multi-faceted. However, this DP may be difficult to understand for some people, mainly due to the perceived challenge of planning multi-faceted metrics. Thus, more guidelines on how to implement it and practically define the best multi-faceted criteria and metrics would be useful, which would require future research.

DP2 was regarded by DM 1 as one of the most important DPs since it is crucial to understand the decision-making process, the roles involved, and the interrelationship of the data-driven decision making elements. Nevertheless, documenting a unified viewpoint in an organization is complicated, and is recommended to be planned out before implementation to make it clear to all decision makers. While the DP itself is easy to understand, and its necessity and value is quite clear, it may be hard to act upon, depending on the organization. Oftentimes, it is challenging during implementation to define and model the structure in a way comprehendible to everyone, particularly those from different backgrounds. Hence, standardized guidelines or structures aiding the modeling would be necessary.

The proper implementation of DP2, supports DP3 and simplifies its realization. In DP3, the concept of collaborative rationality is new, and the modes of collaboration and roles of decision makers need to be defined. For machines and humans to learn together, they need to collaborate together with transparent aims and understand the modes and reasons for collaboration. For example, in Antell’s case, it was noticed that some restaurant managers didn’t always implement ATSO’s decision and may have spontaneously chosen to implement their own decisions when the role of each was not clear. However, in many such scenarios, deviating from ATSO’s decision led to losses (e.g. financial). On the other hand, ATSO was also prone to error, particularly in the early stages, and sometimes made inappropriate suggestions which needed to be modified by humans. Accordingly, it is important to know that decision makers are utilizing the AI system and collaborating with it in the right way, as well as taking advantage of its capabilities to augment their own intelligence and competences. By defining this collaborative rationality, it becomes clear for all parties involved, and the chances for deviation become limited. However, more structure and guidelines on how to accurately define and enhance collaborative rationality would be useful.

DP4 was also deemed important, novel, and effective. Potential difficulty in implementation is due to lack of knowledge of the events and changes that may trigger processual ex-post evaluation in advance. Routine evaluations are easier to set, particularly if an evaluation process or a structure on how to conduct the evaluations is specified.

Moreover, DP5 was also found to be important and novel, and strongly linked to DP4. While the DP itself is easy to read, defining how it will be acted upon may be difficult since there are various kinds of feedback leading to different loops. Additionally, it raises further questions, such as where each feedback loop goes, how it can adapt and evolve, what the iteration length is, how to automate the feedback and learn in retrospective, or perform posterior rationalization after each loop, and how to maximize the effectiveness of the feedback and learning loops. While the DP is still feasible and can be acted upon, it requires thorough planning and guidelines for proper implementation.

5. Discussion and Conclusion

The main contribution of our paper is *proposing a set of five DPs for systems implementing data-driven decision evaluation and evaluating them in a case study which put them into action.* These DPS have been derived from theory and practice and evaluated in the practical case of Antell. In this context-specific case, experienced DMs evaluated their reusability and judged the DPs to be accessible, important, novel, actable, and effective. While theoretical novelty is generally assessed by the academic community during the publication process, practical novelty must be assessed by the experts and users of the DPs [25], which in our case found all the DPs to be novel and formulated in a generalized way that they can potentially be applied and evaluated in other contexts. This highlights the contribution of the DPs to theory and valuable addition to the comprehensive knowledge base as prescriptive statements indicating how to build [10, 11] systems implementing data-driven decision evaluation, and serving as level 2 nascent design theories in DSR which can be used to generalize towards a design theory [12]. Additionally, the DPs contribute to practice by providing theoretically grounded and practically feasible statements to guide developers on how to build systems implementing data-driven decision evaluation to support the aims indicated in each DP.

DP1 extends beyond works such as [26] and [27] which show the limitations and insufficiency of primary performance measures used by managers to evaluate AI tools and their output, as well as the conflict between the reported measures of such tools and the actual decision results, and the knowledge of experts who are capable of considering various important variables [28]. Hence, DP1 provides an implementable solution to these issues by prescribing the use of multi-faceted evaluation criteria and metrics which consider various contexts and aspects to govern the entire data-driven decision. It challenges the use of limited and myopic evaluation metrics (e.g. accuracy, confidence, specificity, sensitivity, area under the curve (AUC), etc.) in current research, and supports that decision outcomes are the ultimate indicators of success and multiple factors need to be considered in evaluation [29, 30].

Moreover, DP1 and DP2 contribute to having a comprehensive viewpoint for the multiple, socio-technical, elements involved in data-driven decision making which is lacking, albeit necessary, in current literature [31]. DP2, which prescribes having a unified viewpoint of the data-driven decision making elements and evaluation, further supports the strategy for data-driven decision making system development, transparency, and integration called for in [3]. Granted that every stakeholder has their own views, the evaluation criteria of DP1 need to be explicitly defined, and the evaluation grounds must be clarified. While the other DPs are in line with the design objectives explicated in [4], DP 2 in particular emerged from practical collaboration in the case of Antell and its development of ATSO when it was realized that different decision makers and stakeholders had different views of the system, the decision, and decision evaluation, which led to complications, biases, and misinformed evaluations. This highlights the importance of design science research and working alongside practice to encapsulate practical requirements into theoretical contributions. Accordingly, DP2 presents an importance concept for multi-stakeholder decision making systems and shows the importance of modelling for enforcing stakeholders to unify the data-driven decision making elements and their understanding of them. Such unification can further support monitoring in metahuman systems [6] by understanding the interrelationship between the elements and how changes in one area affect the system as a whole.

DP3, which emphasizes defining collaborative rationality and the mode of collaboration between humans and machines, builds on the body of literature on human-machine collaboration such as [15] and [14], and prescribes a statement for facilitating such collaboration in practical implementations. Furthermore, acting on the feedback and learning received from evaluation, as depicted in DPs 4 and 5, helps enhance collaborative rationality and human-machine collaboration, and aids in illustrating a joint understanding of the data-driven decision elements. By evaluating the decisions, it becomes easier to understand the aspects in collaboration which add to the quality of the decision and to enhance human-machine collaboration [5]. Thus, DP3 concretizes how to put the concept of delegation in metahuman systems [6] into practice by understanding ex-post what happens with shifts of agency in the system and changes in collaboration modes, and what decision-making tasks to delegate and their consequences.

DP4 adopts a process science perspective [18] for capturing changes in the environment and understanding how the data-driven decision elements evolve, interact, and unfold across time. Current literature is lacking in supporting the long-term follow up necessary in data-driven decision making systems [26], and potentially overlooks the aspect of time and the requirements for a longitudinal, or processual evaluation. However, this is crucial for capturing the complex dynamics involving change and understanding the phenomena that led to the outcomes [18]. DP4 attempts to fill this gap in theory and practice by prescribing processual ex-post evaluation across time.

DP5 supports system development through experiential learning towards more automated system adaptivity [21] by implementing feedback loops that evolve and adapt to the changing requirements and contexts across time, which are realized in DP4. Hence, DP4 and DP5 support reflecting in metahuman systems through a longitudinal analysis for building novel knowledge [6]. Furthermore, DP5 adds to recent research which accentuates the use of a feedback loop for enhancing organizational learning and the learning and utilization of the ML tool (e.g. [2, 32]). However, DP5 contributes beyond such research by enabling not only single-loop or double-loop learning, but also deutero learning, as well as providing prescriptive statements for practical implementation. Thus it enables reflecting through such learning for the continuous evaluation and improvement of human-machine collaboration and data-driven decision making systems [3, 6, 16, 32].

Although the evaluation of the DPs was limited to a single context and case, the DMs agreed that having our proposed DPs structures and simplifies the development of the whole data-driven decision making system or its modularized components, and may be applicable in other contexts. Future research includes formulating more defined guidelines and principles for implementation of the DPs, possibly through the generation of design patterns for system development, as well as monitoring the longitudinal effect of the DPs on enhancing data-driven decision making.

References

[1] Elgendy N, Elragal A, Päivärinta T. DECAS: a modern data-driven decision theory for big data and analytics. *Journal of Decision Systems* 2021; 1–37.

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

[3] Dwivedi YK, Hughes L, Ismagilova E, et al. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management* 2021; 57: 101994.

[4] Elgendy N, Elragal A, Ohenoja M, et al. Ex-Post Evaluation of Data-Driven Decisions: Conceptualizing Design Objectives. In: Nazaruka et al. Ē (ed) *Perspectives in Business Informatics Research*. Springer Nature Switzerland, 2022. Epub ahead of print 2022. DOI: https://doi.org/10.1007/978-3-031-16947-2_2.

[5] Elgendy N, Elragal A, Päivärinta T. Evaluating collaborative rationality-based decisions: a literature review. In: *Procedia Computer Science*. 2023.

[6] Lyytinen K, Nickerson JV, King JL. Metahuman systems = humans + machines that learn. *Journal of Information Technology* 2020; 36: 427–445.

[7] Golovianko M, Gryshko S, Terziyan V, et al. Responsible cognitive digital clones as decision-makers: A design science research study. *European Journal of Information Systems* 2022; 0: 1–23.

[8] Rinta-Kahila T, Someh I, Gillespie N, et al. Algorithmic decision-making and system destructiveness: A case of automatic debt recovery. *European Journal of Information Systems* 2022; 31: 313–338.

[9] Mikalef P, Conboy K, Lundström JE, et al. Thinking responsibly about responsible AI and 'the dark side' of AI. *European Journal of Information Systems* 2022; 31: 257–268.

[10] Chandra L, Seidel S, Gregor S. Prescriptive Knowledge in IS Research: Conceptualizing Design Principles in Terms of Materiality, Action, and Boundary Conditions. In: *2015 48th Hawaii International Conference on System Sciences*. HI, USA: IEEE, pp. 4039–4048.

[11] Gregor S, Kruse L, Seidel S. Research Perspectives: The Anatomy of a Design Principle. *JAIS* 2020; 21: 1622–1652.

[12] Gregor S, Hevner AR. Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly* 2013; 37: 337–355.

[13] Tversky A, Kahneman D. The Framing of Decisions and the Psychology of Choice. *Science* 1981; 211: 453–458.

[14] Ransbotham S, Khodabandeh S, Kiron D, et al. Expanding AI's Impact With Organizational Learning. *MIT Sloan Management Review*, https://sloanreview.mit.edu/projects/expanding-ais-impact-with-organizational-learning/ (2020, accessed 22 December 2021).

[15] Shrestha YR, Ben-Menahem SM, von Krogh G. Organizational Decision-Making Structures in the Age of Artificial Intelligence. *California Management Review* 2019; 61: 66–83.

[16] Duan Y, Edwards JS, Dwivedi YK. Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management* 2019; 48: 63–71.

[17] Rescher N. *Process metaphysics: an introduction to process philosophy*. Albany: State University of New York Press, 1996.

[18] vom Brocke J, van der Aalst W, Grisold T, et al. *Process Science: The Interdisciplinary Study of Continuous Change*. SSRN Scholarly Paper ID 3916817, Rochester, NY: Social Science Research Network. Epub ahead of print 3 September 2021. DOI: 10.2139/ssrn.3916817.

[19] Pettigrew AM. Contextualist Research and the Study of Organizational Change Processes. 1985; 20.

[20] Argyris C, Schön DA. Organizational Learning: A Theory of Action Perspective. *Addison-Wesley* 1997; 345.

[21] Oppermann R, Rashev R. Adaptability and Adaptivity in Learning Systems. *Knowledge Transfer* 1997; 173–179.

[22] Blöcher K, Alt R. AI and robotics in the European restaurant sector: Assessing potentials for process innovation in a high-contact service industry. *Electron Markets* 2021; 31: 529–551.

[23] Lai HBJ, Karim S, Krauss SE, et al. Can restaurant revenue management work with menu analysis? *J Revenue Pricing Manag* 2019; 18: 204–212.

[24] Peters K, Hervé Remaud Pr. Factors influencing consumer menu-item selection in a restaurant context. *Food Quality and Preference* 2020; 82: 103887.

[25] Iivari J, Hansen M, Haj-Bolouri A. A Proposal for Minimum Reusability Evaluation of Design Principles. *European Journal of Information Systems* 2020; 30: 286–303.

[26] Lebovitz S, Levina N, Lifshitz-Assa H. Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What. *MISQ* 2021; 45: 1501–1526.

[27] Ioannidis JPA, Cripps S, Tanner MA. Forecasting for COVID-19 has failed. *Int J Forecast*. Epub ahead of print 25 August 2020. DOI: 10.1016/j.ijforecast.2020.08.004.

[28] Herm-Stapelberg N, Rothlauf F. The crowd against the few: Measuring the impact of expert recommendations. *Decision Support Systems* 2020; 138: 113345.

[29] Magrabi F, Ammenwerth E, McNair JB, et al. Artificial Intelligence in Clinical Decision Support: Challenges for Evaluating AI and Practical Implications: A Position Paper from the IMIA Technology Assessment & Quality Development in Health Informatics Working Group and the EFMI Working Group for Assessment of Health Information Systems. *Yearb Med Inform* 2019; 28: 128–134.

[30] Krupinski EA. Evaluating AI Clinically—It's Not Just ROC AUC! *Radiology* 2021; 298: 47–48.

[31] Namvar M, Intezari A. Wise Data-Driven Decision-Making. In: Dennehy D, Griva A, Pouloudi N, et al. (eds) *Responsible AI and Analytics for an Ethical and Inclusive Digitized Society*. Cham: Springer International Publishing, 2021, pp. 109–119.

[32] Sturm T, Gerlach J, Pumplun L, et al. Coordinating Human and Machine Learning for Effective Organization Learning. *MISQ* 2021; 45: 1581–1602.