Proceedings of the 57th Hawaii International Conference on System Sciences | 2024

Understanding Cognition in the Development of Artificial Intelligence-based Systems: An Exploration of Cognitive Fit and Supporting Mechanisms

Mei Wu-Gehbauer
University of Cologne
m.wu-gehbauer@wiso.uni-koeln.de

Christoph Rosenkranz
University of Cologne
rosenkranz@wiso.uni-koeln.de

Phil Hennel
University of Bremen
hennel@uni-bremen.de

Abstract

Artificial intelligence (AI) increasingly supports users in their work practices in organizations. To unlock value from AI, users need to interact with AI-based systems in their decision-making processes. With increasing sophistication of AI, understanding the mechanisms influencing the cognitive fit between users and AI-based systems is crucial to ensure a successful implementation. Drawing upon literature on cognition in IS and employing a revelatory case study, we explore mechanisms enacted by data science teams in shaping a cognitive fit. We find that the data science teams enact supporting mechanisms to enable users to make sense of AI-based systems and AI "making sense" of users mental models. With this, we contribute to the body of knowledge on cognitive fit in IS by shedding light on underlying mechanisms for an effective development of AI-based systems. For organizations, these insights are valuable to overcome barriers to the successful introduction of AI-based systems.

Keywords: Artificial intelligence, cognitive fit, data science team, supporting mechanisms

1. Introduction

Given the rapid advancements in *artificial intelligence* (AI) and its unique opportunities to boost automation, decision support, and customer experience, organizations continue to invest heavily in AI capabilities. Global spending on AI solutions will reach $154 billion by end of 2023, an increase of almost 30% over the amount spent in 2022 (Insights, 2023). Although there is a strong and increasing interest in AI and companies are investing significantly in it, the reality remains that over two-thirds of AI projects fail to meet expectations (Weiner, 2022). The biggest mistake of organizations is to consider AI-based systems as "plug-and-play" solutions and to expect immediate returns (Fountaine et al., 2019). This overlooks the fact that most obstacles in deriving value from AI are human-related (Bean & Davenport, 2019). With the increasing integration of AI-based systems working alongside and augmenting users instead of replacing them (Jain et al., 2021), it becomes crucial to comprehend how users perceive and interact with these systems.

Users face a great deal of ambiguity in making sense of AI-based systems in their work practices. On the one hand, they need to deal with the unique facets of AI, i.e., self-learning, inscrutability, and autonomy that are different from previous generations of IS technologies (Berente et al., 2021). On the other hand, they face novel challenges in their work practices, such as balancing the practice of human agency and machine autonomy (Lindebaum et al., 2020), being able to give reasons and explain why a decision has been taken by the AI-based system (Lebovitz et al., 2022), or ensuring the ability to learn and maintain unique human knowledge (Fügener et al., 2021). This shifts the locus of action, choice, control, and power away from the exclusive domain of humans (Benbya et al., 2021), requiring an understanding of cognition in the collaboration of humans with the AI-based system (Dafoe et al., 2021; Fulbright & Walters, 2020).

Current research on AI-based systems and their value creation provides only partial answers on how users make sense of AI and often does not consider the role of cognition to support this. Mostly, existing studies on value creation from AI focus on a strategic understanding on an organizational level on required AI capabilities (e.g., Mikalef & Gupta, 2021). Moreover, prior research has predominantly investigated cognition with a focus on the post-deployment phase during the operational use of AI-based systems (e.g., Fügener et al., 2022). While these studies provide valuable insights into the impact of AI-based systems when using the systems, they do not address the specific cognitive processes and decision-making strategies that users employ during the design and development phase. This hinders our understanding of how users' cognitive needs and expectations shape the design and functionality of AI-based systems during the development phases.

URI: https://hdl.handle.net/10125/106474
978-0-9981331-7-1

A prominent IS research stream addressing the perception and interaction of users with IT-based systems is cognitive fit theory (Vessey & Galletta, 1991). Developing an appropriate mental representation (cognitive fit) between technology with users' tasks leads to better decision-making performance, whereas cognitive dissonances can lead to decreased performance (Vessey & Galletta, 1991). We know that understanding individual cognition and the collaboration between users and stakeholders such as developers or systems analysts can lead to better knowledge transfer, mutual trust, and shared mental models, leading to more successful IS outcomes (Chakraborty et al., 2010). Several authors applied a cognitive fit lens in the context of AI, such as for studying the post-adoption use of virtual assistants (Rzepka et al., 2020), cognitive load of system developers (Kam & D'Arcy, 2022), and how AI-based systems can support cognitive fit by adapting information representations (Samuel et al., 2022). However, achieving cognitive fit is far more challenging in the context of humans collaborating with AI-based systems. We identified two key aspects of AI-based systems that limit users' ability to enact conventional fit mechanisms to achieve cognitive fit in the context of AI-based systems. First, the increasing difficulty to make sense of the functioning of the AI-based systems (due to the facets of self-learning and inscrutability), and second, the required alignment with the embedded "mental model" of AI-based systems.

We suggest that the data science team plays a central role in addressing these obstacles for users while developing the AI-based system. Existing research largely neglects the role of the data science team and their interactions with users to make sense of AI in the business context and take decisions based on AI-based insights. Based on this, our study aims to examine the following research question:

"How do data science teams interact with users during development of AI-based systems to support and help them in bridging cognitive dissonances resulting from AI to make sense of AI-based systems in their work environment?"

To answer our research question, we use an exploratory case study of data science projects in a large, multi-national telecommunications company. We focus on the cognition processes of users when interacting with AI, and the role the data science team plays for this. With this, we uncover mechanisms the data science teams enact to address cognitive obstacles users encounter when interacting with AI.

The rest of the paper is structured as follows. First, to investigate the research question, we conceptualize AI and the human actors involved. Then, we draw on cognition theory to shed light on users' cognitive fit and the supporting mechanisms enacted by the data science team. To gain empirical evidence, we conducted an exploratory case study interviewing representatives from the data science team and the users. Finally, we conclude by giving theoretical and practical contributions and an outlook.

2. Conceptual background

2.1. AI-based systems & human actors

In our context, we employ the term "AI" primarily to emphasize its technical conceptualization with its unique facets of *self-learning* (i.e., ability of the AI to self-learn), *inscrutability* (i.e., visibility and transparency into the AI algorithm, level of explainability of the algorithm, and level of interpretability and understandability of the algorithm to humans), and *autonomy* (i.e., capacity of the AI to make autonomous decisions), that differentiates AI from previous technologies (Berente et al., 2021). We define AI-based systems as machines performing human-like cognitive functions such as sensing, perceiving, problem-solving, decision-making and innovation in addressing ever more complex decision-making problems (Benbya et al., 2020; Berente et al., 2021).

To create business value from AI, it is essential but not sufficient to invest in building AI assets and capabilities and make use of them to generate insights (Seddon et al., 2017). To realize value from these insights, they must be turned into decisions, and the decisions again must be turned into actions leading to business value (Seddon et al., 2017; Sharma et al., 2014). This is an evolutionary process that emerges out of an active engagement between the human actors within the existing organizational structures, processes, and context (Seddon et al., 2017; Sharma et al., 2014). Without these human-intensive actions towards implementing AI-based decisions, value creation from AI cannot be realized (Bean & Davenport, 2019).

Human actions are performed by human actors and Davenport (2021) identifies three main groups of human actors in relation to analytical work: (1) the *data science team*, (2) the *decision makers*, and (3) diverse groups of (*end) users*. The data science team is a multi-disciplinary team consisting of data scientists, machine learning engineers, and IT experts, who apply analytical methods to generate unique insights from data to outside groups. The decision makers are the data leaders who set the tone for an analytical culture and make important decisions based on AI insights. The users can be composed of the (end) users, who use the product directly or indirectly, and of user representatives. In large organizations, it is fairly common to have user representatives who are domain experts (i.e., those who not only have an intricate knowledge about the users' business processes but are also somewhat familiar with systems analysis techniques). Typically, in agile deliveries the role of the user representative is assumed by the product owner (PO). To succeed with AI, these distinct groups of human actors need to actively engage with each other.

In this study, we focus on the users / user representatives and the data science teams as well as on the interactions between these two groups of actors involved in developing the AI solution. Data leaders are interviewed for triangulation purposes.

2.2. Cognition in the context of AI

We draw on a cognitive approach to study the interactions between human actors and how people make sense of the AI-based systems they develop and use. A cognitive approach to studying IS in organizations is not new (Tan & Hunter, 2002). Earlier studies showed that, for instance, managerial cognition can lock firms into existing market and technological trajectories resulting in difficulties to adapt to radical technological change (Tripsas & Gavetti, 2017). Along the same lines, Orlikowski and Gash (1994) argue that understanding the cognition of the stakeholders can lead to more successful information systems outcome. Cognition has been used in IS research to better understand how people think about IS in the organization. It is a broad construct. Conceptualization used ranges from "schemas" (Cossette & Audet, 1992), "cognitive maps" (Weick & Bougon, 2001), "technological frames" (Orlikowski & Gash, 1994), to "mental models" (Daniels et al., 1995).

In this research, we build upon the cognitive concept of mental models to describe the underlying assumptions, expectations, and knowledge that individuals or groups have and rely upon when making decisions about IS use in their context. Previous research shows that mentals models are especially important in the development phase of IS (Chakraborty et al., 2010). People draw on their mental models in their effort to make sense of AI-based systems in their work practices and engage in cognitive alignment work. We draw on work from socio-cognitive theory to conceptualize user mental models in the context of AI as follows: (1) Knowledge about the AI-based system (what it is and what it is not), and understanding what it does (Orlikowski & Gash, 1994), i.e. understanding of the main features of AI and of the functioning of the AI-based system, (2) explanations and error detection (Carroll & Olson, 1988), i.e., being able to explain why certain outputs are erroneous and being able to detect and recover from errors from the AI-based system, and (3) cognitive augmentation strategies (von Krogh, 2018), i.e., understanding of how to best use AI to augment human cognitive capabilities in the human-AI collaborative setup. Figure 1 illustrates the conceptualization of the mental model we draw on.

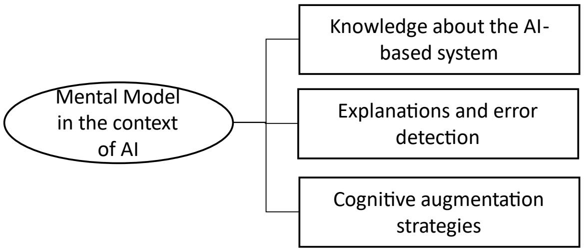

Figure 1. Conceptualization of mental models in the context of AI (integrated from Carroll & Olson, 1988; Orlikowski & Gash, 1994; von Krogh, 2018)

2.3 Cognitive fit for AI-based systems

Cognitive fit theory suggests that a good match between users' cognitive processes and the cognitive demands of the IS leads to better performance and satisfaction (Vessey & Galletta, 1991). However, there are two aspects of AI-based systems that make it increasingly challenging for users to adapt their mental models in a way to align with the cognitive demands of AI-based systems and rely on conventional mechanisms to achieve cognitive fit.

First, it is assumed in literature that the functioning of the system, the relationships of input and output, and how the system arrives at a certain output is knowable to the users and users can make sense of and adapt their mental models based on their user experience to achieve cognitive fit (Zhang et al., 2021). For AI-based systems this is increasingly difficult to achieve due to their generativity resulting from self-learning, i.e., the technology's ability to produce unprompted change (Yoo et al., 2012), the complexity and opacity of the algorithms (Burrell, 2016; von Krogh, 2018), and their ability to take actions autonomously without human interaction or delegation (Berente et al., 2021). Moreover, predictions about future actions of the AI-based system based on past experiences with the system are increasingly difficult to make due to its high variability. Zhang et al. (2021) show that the outputs of AI-based systems are in a state of continuous emergence, and the interactions of the users with the systems are marked by ambiguity and surprises.

Second, in contrast to conventional IS technology that supports users in task completion, the AI-based system has, just as humans, an own embedded "mental model" in form of algorithms that augment the cognitive capabilities of users in a collaborative human-AI setup (Seidel et al., 2018). These mental models change over time and can be applied to a variety of multiple tasks and problem domains (Berente et al., 2021; Seidel et al., 2018). Therefore, cognitive fit in this context goes beyond developing an appropriate mental representation of the technology with a task at hand. In a human-AI collaborative setup, users need to understand how their own mental model interacts with the mental model of the AI-based system in a variety of problem domains, which is fundamentally distinct from conventional IS technology (Grimes et al., 2021). Such an understanding is required, for instance, to effectively distribute tasks between users and the AI (Fügener et al., 2022), and to enable users to decide when to override the automated inference (Bansal et al., 2019). Given the increasing level of cognitive augmentation by AI-based systems towards higher level cognition – i.e., concept understanding, creative human inspiration, and conversational natural language (Fulbright & Walters, 2020) and the continuing expanding frontier of AI (Berente et al., 2021) – developing and maintaining an accurate mental representation of the systems capabilities as well as the boundaries of the AI-based system are increasingly difficult (Bansal et al., 2019; Grimes et al., 2021).

As a result, users are limited in their ability to refine their mental models to achieve cognitive fit that just relies on conventional cognitive "fit" mechanisms, such as designing the system in a way that is compatible with pre-defined user tasks and work processes (Goodhue & Thompson, 1995), being able to learn by knowing the input and output relationships of the technologies (Vandenbosch & Higgins, 1996), or providing adequate information representations that match the cognitive needs of the users (Dennis & Carte, 1998). Given the lack of effectiveness of the conventional "fit" mechanisms, we argue that for users it is often even impossible to adapt their mental model in a way to align with the AI-based system on their own. We suggest that the data science team plays a central role in addressing these obstacles for users while developing the AI-based system.

3. Research method

We employ an exploratory case study design for our study, since user cognition in the context of developing an AI-based system is a phenomenon that is not well researched and understood. At the same time, it is growing in importance due to the increasing cognitive obstacles users are facing with the rapid advancements and increasing complexity of AI technologies. An empirical, exploratory study is best suited to investigate this in an in-depth manner within the real-life context of an organization. To address the outlined research question, we study a revelatory single case (Dubé & Paré, 2003). A single case study allows for developing an in-depth understanding on cognition in the context of IS based on empirical account (e.g., Davidson, 2002). We study several AI projects from different perspectives on the individual level and the team level. For the single, embedded design, we follow established guidelines for conducting an exploratory case study research (Eisenhardt, 1989; Yin, 2009).

3.1. Research setting

Our case organization is a leading multi-national telecommunications provider and will be referred to as TelcoGroup hereafter. TelcoGroup serves both private customers with broadband, mobile, and Internet-based media products and services as well as business customers with information and communications technology (ICT) solutions. Over the last years, TelcoGroup has invested extensively in AI capabilities. It deploys data science teams applying a wide range of different AI technology sophistication levels. TelcoGroup was chosen as a revelatory case organization because it presents a unique opportunity to investigate our research question, with extensive access to internal data and teams across different business and IT areas, as well as the possibility to collect longitudinal data. The main units of analysis are individuals and teams involved in developing AI-based systems. For this study, data was collected from several teams as well as users, who all operate within the case organization.

3.2. Data collection and analysis

Interviews were the main form of gathering data, next to presentations, documentations, and press-releases. The selection of interviewees followed a combination of literal (conditions of the cases lead to predicting the same results) and theoretical (conditions of the cases lead to predicting contrasting results) replication logics to analyze the different setups in which the interviewees operate (Dubé & Paré, 2003). We conducted 28 semi-structured interviews with data science team members and users from eight different AI projects. The projects cover AI-based use cases for predictive and prescriptive maintenance and optimization of network and equipment, chatbots in helpdesk, autonomous agents in general management and administrative functions, predictive ticket routing in customer support, and personalization of customer recommendation and communication applying machine learning including deep learning and neural networks. The interviews are based on an interview guide including questions about the interviewees' role, and experience in the organization, about the AI-based system in development, the knowledge, attitude, and beliefs of the interviewees towards AI, as well as the interactions between the data science team and the users.

The interviews were usually conducted in a group of two to three researchers. Additionally, a questionnaire to capture the key characteristics of the AI-based systems and the user involvement has been completed by interviewees following the interview. All interviews were recorded with prior consent by the interviewee and transcribed to ensure all information is captured for data analysis. Table 1 provides an overview of the interview data.

Table 1. Interview data

Position Interviewee	Number of interviews	Interview time in minutes
Data science team member (DS)	13	695min
User / user representative (User)	11	587min
Data leader (DL)	4	210min
Total	*28*	*1.492min*

The data was coded following a form of grounded theorizing (Glaser & Strauss, 1967). Following an initial round of interviews, the data was analyzed building on the grounded theory data analysis methods of open, selective, and theoretical coding (Glaser, 1978; Saldaña, 2009).

As coding can be seen as a "cyclical act" (Saldaña, 2009), our coding process therefore was very iterative and formed a mutually influencing relationship with the process of data collection. Based on the codes building on extant literature (e.g., those building upon mental models and cognition) and later the emerging codes, we set out to identify additional constructs and refine constructs by means of pattern coding as described by Miles and Huberman (1994) and Saldaña (2009). Pattern coding is appropriate for the development of major themes from data (Miles & Huberman, 1994; Saldaña, 2009). These codes are capable to "identify an emergent theme" and therefore are helpful for "grouping those summaries into a smaller number of sets, themes, or constructs" (Miles & Huberman, 1994, p. 69). In addition to the transcripts, the coding was also triangulated using supplemental material (e.g., field notes, instructional material / managerial guidelines) to increase validity. Furthering our understanding of the concepts and their interactions, we coded relationships between them and augmented our data by including additional sources here as well. Based on these steps and through iterative coding, discussion, revision, we finally converged to a shared understanding and agreed-upon theoretical explanation for the observed phenomena.

We followed three tactics to increase construct validity (Lee, 1989). We used multiple sources of evidence (multiple key informants and data sources) and established a chain of evidence (case study database) during data collection. Furthermore, key informants reviewed draft reports of the case study. In the data analysis, we addressed internal validity by pattern matching (linking the resulting concepts and interactions to data from the case study database) and explicit explanation building.

4. Findings

The findings from our case study support that users suffer cognitive dissonances when developing AI solutions. User and data science teams have different but complementary mental models. While the data science teams bring in their expertise in analyzing large amounts of data and in developing AI-based systems, the users bring in their domain knowledge and experiences from their work practices. We found that beyond aligning their cognitive representations, the data science team engaged in supporting mechanisms to bridge the cognitive gap between the mental model of the user and the mental model of the AI-based system. We identify two types of supporting mechanisms that the data science team enacts: (1) mechanisms supporting the *users making sense of the mental model of the AI-based system*, and (2) mechanisms supporting the *AI-based system "making sense" of the mental model of the users*.

We first present the user cognitive barriers in developing AI solution. Then we outline the supporting mechanisms enacted by the data science team.

4.1. User cognition barriers in developing AI-based systems

In the AI projects we studied, we found that users suffered from cognitive dissonances with the AI-based system due to (1) the unique characteristics of the AI-based system of uncertainty, variability and inscrutability imposing barriers for users to make sense of the functioning and outputs of AI, and (2) missing cognitive augmentation strategies in a human-AI collaborative setup. In the following section, we describe these barriers with exemplary quotes.

In the first place, users have problems grasping the specific facets of AI of uncertainty, variability, and inscrutability. Outcomes of AI include probabilistic predictions, and randomness to explore different strategies or generate diverse outputs and recommendations. This is very different from rule-based and deterministic outcomes of traditional IT systems. The following quote illustrates this challenge the users are facing: *“These models will not make predictions that are 100 percent accurate. But I remember back at the beginning when people used to say, ‘You just need to keep training it, and eventually, we'll reach 100 percent accuracy, and everything will be great.’”* (DS, interview 8). In addition, users show discomfort when confronted with the inscrutability of AI-based systems. They express the need to understand how outputs are being generated, as the following quotes demonstrate: *“People are concerned about using the AI solution. […] we have often encountered the issue of the ‘blackbox neural network’”* (User, interview 7). *“For me, it is indeed important to understand how the algorithm generates the outputs. If I have a result and have no idea how that result came about, I can't interpret it properly. I don't know which levers to adjust to get the right result.”* (User, interview 26).

Next, users struggle with identifying strategies for cognitive augmentation in a joint human and AI collaboration setup. Challenges we found involve users being frustrated about perceiving a loss when applying AI and trying to make sense of how a collaboration with AI can add value. The following quotes reflect the challenges of the users: *“Our colleagues were always striving to find the perfect solution. Just a year or two ago, we couldn't accept that there are also solutions that are only 80 percent accurate and 20 percent incorrect. But if the AI can handle 10,000 issues, and out of those 10,000, 8,000 are top-notch, still a lot more have been processed well compared to previously. It always depends on the overall quantity that gets processed. On what turns out positively, even if we have some loss. Only this loss challenges the ‘optimization pride’ that our colleagues hold.”*(User, interview 28). “*We no longer have that deep granularity of individual cases in our daily routine. With AI, we're operating at a higher altitude, and there's a certain level of fuzziness in it. We don't catch every single issue anymore. On the other hand, in terms of workload, we have to set some boundaries. We work based on priorities that were automated through an algorithm. We follow those priorities from the top down. And this bothers many people because they no longer have that freedom.”* (User, interview 22).

4.2. Supporting mechanisms

We identified two types of supporting mechanisms enacted by the data science team to address users' cognitive barriers towards AI: (1) *user making sense of the AI mental model*, and (2) *AI “making sense” of the user mental model*.

User making sense of the AI mental model. The data science team actively enacted the mechanism to support *users making sense of the AI mental model* referring to helping users to adjust their mental model to accommodate the uncertainty, variability and inscrutability associated with AI-generated insights. Building on and in addition to supporting the users to resolve the cognitive dissonances, the data science team plays a key role in facilitating the users in making sense of how to collaborate with the AI-based system in a human-AI collaborative setup.

For users to make sense of the AI mental model, we found four supporting mechanisms performed by the data science team which we detail in the following section with exemplary quotes.

Translating into business impact. The translation into business impact provides users with an overarching understanding on the rationale of applying the AI-based system. This includes the mechanism of translating the performance KPIs of the AI-based system into how they manifest in terms of business value: *“What does it mean to me if I have an accuracy of 70 percent - how much benefit do I get. […] It's about actually calculating the benefit, to say, okay, we have this accuracy here with so many false positives. These false positives incur this much cost, but overall, we still have a positive effect.”* (DS, interview 8). Furthermore, the data science team supports user understanding by translating technical language into the language and logic of the user. “*In the marketing department, it's already a completely different philosophy. You have to translate it because colleagues here deal with entirely different topics in their daily routines. People are primarily focused on commercial matters. So, when data scientists talk in terms of products and customer perception, that works quite well.”* (User, interview 24).

Explaining the outputs of AI. The data science team engages in explaining the outputs of the AI by using feature importance analysis, single-case reasoning as well as simulating the model behavior to explain the outputs of the AI to the users. Feature importance analysis provides the users with the most important features or combination of features that drive the outcomes: *“We've created a table for them in which you can see how important each feature is for making decisions or predictions.”* (DS, interview 15). Single-case reasoning enables the users to investigate single cases to understand the reasoning of the algorithm behind a particular AI system's output: *“They can actually then take a closer look into individual cases and say: Okay, based on their experience, what happened there?”* (DS, interview 8). With simulations, the data science team reviews each run of the algorithm jointly with the users to stepwise understand the emergence of the outcomes and observe how features, weights and rules contribute to the final output: *“We somewhat said, 'If we play around with the individual attributes and* *their values a bit, it's possible that a data point might fall under a different rule or be excluded from a rule.' We examined and found that some attributes tend to co-occur more strongly or have an anti-correlation. We then simply drew a few thousand samples and looked at how it's shaping up now. Under which rules does the data point fall? How likely is it to fall in or out? It was also quite interesting for the users to see that these aren't hard numbers, but rather, everything must be seen through a gentle eye."* (DS, interview 15).

Enabling validation by user. For purposes of validating the results of AI-based system with user expectations, the data science team initiates reviews of the outputs by experts, and provides means to monitor the performance of the AI-based system. Review of the outputs by users aims at aligning the results of the AI-based system with the experience of the users. *"Every two weeks, outputs are being validated by users. It helps to ensure that what the user initially intended aligns closely with the actual results."* (DL, interview 3). Built-in monitoring of an AI system's performance can assist users in validating the impact of the AI-based system and build trust in the system. *"Now we are in the process of teaching the system what it should do. We also see the results of it. We had validation loops built into the system, where we can see what has been worked on and its effect. Did one area improve, or did nothing change, or did it possibly even get worse? We now have a comparison and can see that the results the AI system brings are not so bad. It's in our hands, and we can make it even better."* (User, interview 28).

Transforming AI insights into actions. Beyond building AI-based systems and generating insights from AI, we found that the data science team actively contributes to supporting the users to derive actions from the AI insights. The data science team achieves this by conducting field tests as well as by acquiring business knowledge, allowing them to independently formulate action recommendations. With field tests, user experiences with the AI-based system can be evaluated in the real-world setting helping users to make sense of how to effectively interact with the AI-based system. *"We are currently at an exciting juncture, considering whether and what interactions should occur with the user. Should they receive a notification? If so, how should the notification be worded? There are various directions to explore, ranging from "you must do this" to "take a look and see if not", all the way to automating the process. Right now, a field test is being conducted to determine this."* (DS, interview 8). By gaining business knowledge, the data science team is able to interpret the insights from the AI-based system and in addition to derive value-generating actions for the business on its own. *"The combination of data science knowledge with business knowledge is crucial for interpreting the data. Without these interpretations, the conclusions drawn can be entirely incorrect. Business knowhow is also highly important to derive actions. If a data science team generates results and someone else is responsible for what happens with those results, often, nothing happens at all. You must instill the awareness among data scientists that it's also their responsibility to interpret the results and think about how to turn the results into actions."* (User, interview 24).

AI "making sense" of the user mental model. The data science team enacted the mechanism of *AI "making sense" of the user mental model* referring to feeding back user knowledge and user contexts into the AI-based system, such as unique expert knowhow, preferences, user choices and expectations. This includes revealing user mental models and integrating them in the AI-based system. We found three supporting mechanisms enacted by the data science team that we detail in the following section with exemplary quotes.

User feedback looped back into AI. The data science team establishes feedback loops where users verifies the AI system's outputs based on user expert knowledge and experience. This helps in continuous learning and refinement of the AI-based system and can be done by various means, such as A/B testing. A/B testing allows users to compare the performance of different AI algorithms or AI system configurations to determine which aligns better with their expectations. *"In predictive ticket routing, we actually conduct A/B tests, where we randomly assign users into two groups to determine whether the solution leads to a change in the ticket conversion rate. It's often the case that we initially introduced a solution based on the training data we had seen, believing it would be beneficial. However, it required some fine-tuning before it actually showed a positive effect."* (DS, interview 8).

Output delimitation. The data science team takes measures to limit or enhance the AI output to better fit with user needs and to build in user knowhow into the AI-based system. This can include enlarging the output with additional data to enable more decision options by the users, as well as constraining the output to better control the outcomes and avoid negative user effects: *The first constraint that we imposed into this system, is the pool of possible answers. We don't want [...] a functionality we cannot control and we cannot provide. Therefore, first of all, it's a restricted pool of possibilities."* (DS, interview 3).

Delegation back to the user. The data science team can specify the AI-based system to delegate tasks back to the user, for example for exceptional cases or outliers that are outside of the usual parameters. *"In cases where the use case is highly specific, the expertise and insights from domain experts become increasingly valuable.*

These experts possess an intuitive understanding and a wealth of experience in solving problems using human intelligence, in contrast to automated processes. Their input guides decisions in these particular scenarios, ensuring that the model aligns with what truly matters to the customer. For these specified cases, the AI-based system refers to the human decision." (DS, interview 3).

5. Discussion

The results of our study provide insights into the importance of mutually making sense of the mental models of the users and the AI-based system, as well as the prominent role of the data science team in enacting supporting mechanisms to enable this mutual sensemaking. Overall, we found that given the unique characteristics of AI, users suffer from cognitive dissonances, which cannot be resolved by the users alone due to the unique characteristics of AI. Users' mental models do not fully capture the cognitive demands of AI solutions due to the complexity involved and the (methodological and theoretical) novelty of such systems. While usually such situations call for shared mental models between data science teams and users to be introduced, thereby extending the users' mental models, we found that with AI, users need to make sense of the AI mental model. On the other hand, AI has to consider the user mental models in order for the system to be effectively used.

Based on our findings we identify the following propositions with the key insights of the importance of mutual sensemaking and the prominent role of the data science teams enacting supporting mechanisms for coupling the mental models to elevate the mutual cognitive capabilities.

Mutual sensemaking to elevate cognitive capabilities. The analysis of the cognitive fit in the context of AI-based systems show that the mental models of users and the AI are fundamentally different but complementary. While the AI's ability to process and reason future outcomes from vast amounts of data is unparalleled, user's mental models are built on years of experience and rich with context and nuance. The coupling of user mental models and AI can elevate their joint cognitive capabilities. However, we found that mutual sensemaking is crucial to be able to effectively couple the mental models of the users and the AI. With mutual sensemaking we refer to the users and the AI making sense of their mental models' capabilities and limitations in a way that they come into alignment and stabilize on a higher level of mutual cognitive capabilities. As a result, this may increase trust in the cognitive capabilities of AI-based systems and improve the users willingness to take advice, and act on it, as well as see the technology as helpful, competent, or useful. Users need to be aware that sensemaking of the AI mental model is crucial to develop, in order to elevate their joint cognitive capabilities.

Role of the data science team to enable mutual sensemaking. Users often experience cognitive challenges due to the uncertainty, variability, and inscrutability of AI-generated insights, hindering their ability to make sense of them. Traditional "fit" mechanisms are no longer adequate to align user cognition with AI systems. We found that, given AI's unique characteristics, users struggle to comprehend the AI's underlying mental model. In this context, the data science team plays a pivotal role in facilitating sensemaking in both directions. They implement mechanisms to bridge the gap between user mental models and the AI's mental model, fostering cognitive alignment and enhancing collective cognitive capabilities. This alignment involves users adapting their mental models to better align with the AI system while also enabling the AI to incorporate user knowledge. This approach addresses cognitive challenges related to the uncertainty, variability, and inscrutability of AI-generated insights. By coupling mental models – where users understand the AI system better, and the AI gains insight into user mental models – the cognitive barriers are being tackled.

These propositions are based on the insight from a revelatory case study drawing on the specific lens of cognition and AI solution development. In the following, we outline the implications, limitations, and avenues for future research.

6. Implications, limitations, and future research

In essence, this research contributes to research in the following ways. First, we provide evidence for the importance of user cognition for creating value from AI-based systems. This goes beyond the longstanding insights into the importance of AI capabilities, but rather shows the peculiar qualities of AI solutions and the necessary cognition processes at play. Second, we show how important the data science team is in making sense of these cognition processes and what role it can play in shaping these. This also highlights the idiosyncrasy of AI solutions compared to more traditional information system projects due to the distinct challenges for understanding and working with AI. Third, we introduce and demonstrate the usefulness of cognition and related processes for the discussion on AI in the domain of information system management. Similar to the socio-technical foundations of the information system management discipline, cognition as a socio- and psychological phenomena plays an important role in (successful) AI augmentation strategies.

In terms of practical contribution, emphasizing the importance of users' sensemaking of AI-based systems and recognizing the central role of the data science team in supporting this process is crucial for a human-centered approach to AI. Making sense of the AI mental model as well as the users' mental models as outlined in this paper, and utilizing the supportive mechanisms fully is key to successful AI-based systems development. Given the rise of AI-based systems and the increasingly larger impact of AI on businesses and processes, optimizing the development of such systems will become inevitable.

Naturally, this research is not without limitations. First, while a single case study design provides great depth and focus, it also provides less abstraction from the case organization compared to a multiple case study design. We believe this limitation to be necessary to be aware of but nonetheless to also be of only marginal nature, as this study is of revelatory nature and only the beginning of a larger research project. Further, a widened theoretical lens, especially including institutional theory with concepts such as institutional logics and identity, appears promising. As these concepts play a vital role in team dynamics and organizational processes, and our observations providing initial evidence for such aspects to play a heightened role for AI solutions, we see added value in pursuing this avenue as well.

This research paper presents a first step in conceptualizing the different dimensions of cognition at play in AI-based systems development and provides a distillate of barriers and supportive mechanisms, resulting in a practicable overview for improving cognitive fit of AI-based systems.

7. References

Bansal, G., Nushi, B., Kamar, E., Lasecki, W. S., Weld, D. S., & Horvitz, E. (2019). Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, *7*(1), pp. 2-11.

Bean, R., & Davenport, T. H. (2019). Companies are failing in their efforts to become data-driven. *Harvard Business Review*.

Benbya, H., Davenport, T. H., & Pachidi, S. (2020). Artificial intelligence in organizations: current state and future opportunities. *MIS Quarterly Executive*, *19*(4).

Benbya, H., Pachidi, S., & Jarvenpaa, S. (2021). Special issue editorial: artificial intelligence in organizations: implications for information systems research. *Journal of the Association for Information Systems*, *22*(2), pp. 281-303.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing Artificial Intelligence. *MIS Quarterly*, *45*(3), pp. 1433-1450.

Burrell, J. (2016). How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data & Society 3*(1), pp. 1-12.

Carroll, J. M., & Olson, J. R. (1988). Mental Models in Human-Computer Interaction. *Handbook of Human-Computer Interaction*, pp. 45-65.

Chakraborty, S., Sarker, S., & Sarker, S. (2010). An Exploration into the Process of Requirements Elicitation: A Grounded Approach. *Journal of the Association for Information Systems*, *11*(4), pp. 212-249.

Cossette, P., & Audet, M. (1992). Mapping of an Idiosyncratic Schema. *Journal of Management Studies*, *29*(3), pp. 325-347.

Dafoe, A., Bachrach, Y., Hadfield, G., Horvitz, E., Larson, K., & Graepel, T. (2021). Cooperative AI: Machines Must Learn to Find Common Ground. *Nature*, *593*(7857), pp. 33-36.

Daniels, K., De Chernatony, L., & Johnson, G. (1995). Validating a method for mapping managers' mental models of competitive industry structures. *Human Relations*, *48*(9), pp. 975-991.

Davenport, T. H. (2021). The analytics team. In *INFORMS Analytics Body of Knowledge* (pp. 49-76). Wiley.

Davidson, E. J. (2002). Technology Frames and Framing: A Socio-Cognitive Investigation of Requirements Determination. *MIS Quarterly*, pp. 329-358.

Dennis, A. R., & Carte, T. A. (1998). Using geographical information systems for decision making: Extending cognitive fit theory to map-based presentations. *Information Systems Research*, *9*(2), pp. 194-203.

Dubé, L., & Paré, G. (2003). Rigor in information systems positivist case research: current practices, trends, and recommendations. *MIS Quarterly*, pp. 597-636.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), pp. 532-550.

Fountaine, T., McCarthy, B., & Saleh, T. (2019). Building the AI-powered organization. *Harvard Business Review*, *97*(4), pp. 62-73.

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Will Humans-in-the-Loop Become Borgs? Merits and Pitfalls of Working with AI. *MIS Quarterly*, *45*(3), pp. 1527-1556.

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2022). Cognitive challenges in human–artificial intelligence collaboration: Investigating the path toward productive delegation. *Information Systems Research*, *33*(2), pp. 678-696.

Fulbright, R., & Walters, G. (2020). Synthetic expertise. *Proceedings of the International Conference on Human-Computer Interaction*, pp. 27-48.

Glaser, & Strauss, A. (1967). The Discovery of Grounded Theory: Strategies for Qualitative Research. *New York, Adline de Gruyter*, *17*(4).

Glaser, B. G. (1978). *Theoretical Sensitivity*. University of California.

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, pp. 213-236.

Grimes, G. M., Schuetzler, R. M., & Giboney, J. S. (2021). Mental models and expectation violations in conversational AI interactions. *Decision Support Systems*, *144:113515*.

Insights, F. B. (2023). Big Data Analytics Market Size, Trends | Growth Statistics, 2029.

Jain, H., Padmanabhan, B., Pavlou, P. A., & Raghu, T. (2021). Editorial for the special section on humans, algorithms, and augmented intelligence: The future of work, organizations, and society. *Information Systems Research*, *32*(3), pp. 675-687.

Kam, H.-J., & D'Arcy, J. (2022). A role theory perspective: Will shifting left become a pain for application developers? WISP 2022 Proceedings. 12., Copenhagen.

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. *Organization Science*, *33*(1), pp. 126-148.

Lee, A. S. (1989). Case studies as natural experiments. *Human Relations*, *42*(2), pp. 117-137.

Lindebaum, D., Vesa, M., & Den Hond, F. (2020). Insights from 'the machine stops' to better understand rational assumptions in algorithmic decision making and its implications for organizations. *Academy of Management Review*, *45*(1), pp. 247-263.

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management*, *58*(3), pp. 1-20.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. Sage.

Orlikowski, W. J., & Gash, D. C. (1994). Technological frames: making sense of information technology in organizations. *ACM Transactions on Information Systems (TOIS)*, *12*(2), pp. 174-207.

Rzepka, C., Berger, B., & Hess, T. (2020). Is it a Match? Examining the Fit Between Conversational Interaction Modalities and Task Characteristics. ICIS 2020 Proceedings, Hyderabad.

Saldaña, J. (2009). *The Qualitative Coding Manual*. Thousand Oaks.

Samuel, J., Kashyap, R., Samuel, Y., & Pelaez, A. (2022). Adaptive cognitive fit: Artificial intelligence augmented management of information facets and representations. *International Journal of Information Management*, *65*, 102505.

Seddon, P. B., Constantinidis, D., Tamm, T., & Dod, H. (2017). How does business analytics contribute to business value? *Information Systems Journal*, *27*(3), pp. 237-269.

Seidel, S., Berente, N., Lindberg, A., Lyytinen, K., & Nickerson, J. V. (2018). Autonomous tools and design: a triple-loop approach to human-machine learning. *Communications of the ACM*, *62*(1), pp. 50-57.

Sharma, R., Mithas, S., & Kankanhalli, A. (2014). Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems*, *23*(4), pp. 433-441.

Tan, F. B., & Hunter, M. G. (2002). The repertory grid technique: A method for the study of cognition in information systems. *MIS Quarterly*, pp. 39-57.

Tripsas, M., & Gavetti, G. (2017). Capabilities, cognition, and inertia: Evidence from digital imaging. *The SMS Blackwell Handbook of Organizational Capabilities*, pp. 393-412.

Vandenbosch, B., & Higgins, C. (1996). Information acquisition and mental models: An investigation into the relationship between behaviour and learning. *Information Systems Research*, *7*(2), pp. 198-214.

Vessey, I., & Galletta, D. (1991). Cognitive fit: An empirical study of information acquisition. *Information Systems Research*, *2*(1), pp. 63-84.

von Krogh, G. (2018). Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. *Academy of Management Discoveries*, *4*(4), pp. 404-409.

Weick, K., & Bougon, M. (2001). Organizations as Cognitive Maps: Charting Ways to Success and Failure In KE Weick. *Making Sense of the Organization*, pp. 308-329.

Weiner, J. (2022). *Why AI/data science projects fail: how to avoid project pitfalls*. Springer Nature.

Yin, R. K. (2009). *Case study research: Design and methods* (Vol. 5). Sage Publications.

Yoo, Y., Boland Jr, R. J., Lyytinen, K., & Majchrzak, A. (2012). Organizing for innovation in the digitized world. *Organization Science*, *23*(5), pp. 1398-1408.

Zhang, Z., Yoo, Y., Lyytinen, K., & Lindberg, A. (2021). The unknowability of autonomous tools and the liminal experience of their use. *Information Systems Research*, *32*(4), pp. 1192-1213.