Harmony in autonomy: Navigating the collaboration between humans and autonomous AI agents

Dissertation

zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaft

der Rechts- und Wirtschaftswissenschaftlichen Fakultät

der Universität Bayreuth

Vorgelegt

von

Luis Lämmermann

aus

Roth

Dekan:	Prof. Dr. Claas Christian Germelmann
Erstberichterstatter:	Prof. Dr. Nils Urbach
Zweitberichterstatter:	Prof. Dr. Niklas Kühl
Tag der mündlichen Prüfung:	18.02.2025

If I have seen further, it is by standing on the shoulders of giants.

Isaac Newton

Copyright statement

The following sections are partly comprised of content from research articles included in this thesis. To improve the readability of the text, I omit the standard labeling of citations at these points.

Table of contents

Abstract

Artificial intelligence (AI) is a moving target constantly expanding its frontiers of scope and performance through its increasing intelligence, autonomy, and opaqueness. The growing autonomy drives the paradigm shift in the relationship between humans and information systems. AI artifacts are becoming agentic and are no longer acting on behalf of humans. Instead, AI artifacts are ascending on equal footing with humans and are able to complete tasks without human intervention. From an information systems perspective, the phenomenon of increasing AI autonomy poses significant implications for the collaboration between humans and AI-enabled information systems. This dissertation aims to theorize the management and design of collaboration between humans and autonomous AI agents and guide organizations in successfully implementing AI endeavors along their digital transformation.

I structure my dissertation along three research goals: First, I aim to provide an understanding of managing and designing AI artifacts from an organizational perspective. Second, I aim to foster an understanding of the effective design of human-AI collaboration from an interaction perspective. Third, I aim to theorize the consequences of AI-led interactions between humans and autonomous AI agents from a delegation perspective. To approach the research goals, my dissertation consists of six research essays. Essays 1 to 3 approach my first research goal. Essay 1 theorizes AI application management from an organizational perspective and proposes an AI management model that illustrates AI management's information flows and managerial factors. Essay 2 proposes practices and measures for successfully monitoring machine learning applications. Essay 3 provides a structured systematization of AI platforms guiding organizations in the design of their AI services. Approaching my second research goal, Essay 4 enhances the theoretical understanding of human-AI interaction and proposes design principles for optimal AI advice that fosters human-AI complementarity. To achieve my third research goal, Essay 5 and Essay 6 present phenomena, changing roles, and conflicts arising in this AI-led delegation, contributing to understanding autonomous AI and its implications for human-AI collaboration.

My dissertation contributes to a better understanding of the consequences of the expanding AI frontiers through a pluralistic research approach providing novel theoretical and practical insights into the collaboration between humans and autonomous AI agents.

Keywords: artificial intelligence, autonomy, management, collaboration, delegation

Introduction to
Harmony in autonomy: Navigating the collaboration between humans and autonomous AI agents

Abstract

This dissertation seeks to illuminate the design and management of collaboration between humans and autonomous artificial intelligence (AI) agents within organizational settings. It consists of six essays, which have either been published in or submitted to renowned journals and conferences. The essays collectively address how organizations can effectively manage and design AI artifacts, establish optimal human-AI collaboration, and understand the consequences of AI-led interactions between humans and autonomous AI agents from a delegation perspective.

The introduction of my dissertation comprises six sections. In Section 1, I motivate the overall relevance of my research. In Section 2, I provide the relevant theoretical background of my research. In Section 3, I identify research gaps and formulate questions around three research goals. Section 4 outlines the overall research design and methodological approach. In section 5, I summarize the results of my research. Lastly, section 6 concludes the dissertation by discussing the results, addressing the study’s limitations, and suggesting directions for future research.

Keywords: artificial intelligence, autonomy, management, collaboration, delegation

Table of contents

1 Motivation

Artificial intelligence (AI) agents are increasingly pervading today's society, providing businesses and individuals with unprecedented opportunities (Berente et al., 2021; Candrian & Scherer, 2022). Businesses incorporate AI agents to enhance operational efficiency through advanced process automation, while also integrating AI into products and services enabling new business models for customers (Agrawal et al., 2019; Davenport & Ronanki, 2018; Huang & Rust, 2021). Individuals interact with AI agents for levering numerous tasks, from financial management and personalized healthcare support to facilitating work routines (Puntoni et al., 2021; Sauerbrei et al., 2023; Zhu et al., 2024).

AI's recent pervasion is the result of decades of research efforts relating back to the 1950s (McCarthy et al., 1955). Since then, AI was a scientific niche topic for long and real-world applications were scarce, mainly owing to limitations in computation power (Benbya et al., 2020). Recent advancements in computational capabilities have led to the rise of AI, enabling realization of previously theoretical technological concepts such as machine learning (ML) at large scale (Berente et al., 2021). The ML concept relies on implicit computational learning from data (Kühl et al., 2022), differing from previous concepts, that relied on explicitly programmed machine behavior. This novel approach allows for remarkable capabilities concerning intelligence, however, at the cost of understanding the inner working and reasoning of such systems (Z. Zhang et al., 2021).

Driven by the expanding intelligence, AI systems are becoming increasingly agentic and autonomous concerning their actions (Berente et al., 2021; Murray et al., 2021; Stelmaszak et al., 2024). Accordingly, AI systems are "*no longer passive tools waiting to be used […] and can now assume responsibility for tasks with ambiguous requirements and for seeking optimal outcomes under uncertainty* (Baird & Maruping, 2021, p. 315). From an IS perspective, AI systems alter the cornerstones of humans' recognition and relationship to information technology (Berente et al., 2021; Lyytinen et al., 2021; Schmitt et al., 2023). AI systems are becoming agents on equal footing with humans exhibiting effective states, delegating and supervising ever-complex tasks (Dennis et al., 2023). Such AI agents also become capable of learning from their own actions and adaptively improving their projective capabilities (Lyytinen et al., 2021). Consequently, the AI agents' increased agency enables more advanced forms of collaboration expressing in novel socio-technical systems (Lyytinen et al., 2021).

Against this backdrop, the pervasion of autonomous AI agents raises new phenomena challenging our existing theoretical understanding in information systems research (Baird & Maruping, 2021; Z. Zhang et al., 2021). Berente et al. (2021, p. 1440) recognize that "*the interaction between humans and autonomous AI is perhaps the key managerial issue of our time"*, calling for dedicated theory and practical approaches to navigate these interactions effectively. Particularly, the ever-expanding facets of AI—namely autonomy, intelligence, and opacity—raise various concerns about trust, performance, and control of such human-AI collaboration (Berente et al., 2021; Ning et al., 2024; Stelmaszak et al., 2024). Consequently, organizations must ensure that human-AI systems foster complementarity while reducing uncertainty in human-AI interactions (Baird & Maruping, 2021; Inkpen et al., 2023; Murray et al., 2021).

Given the practical relevance and the theoretical gaps, the overarching research goal of my dissertation is to shed light on the *design and management of the collaboration between humans and autonomous AI agents*.

I structure the remainder of the introduction as follows: In Section 2, I provide the relevant theoretical foundations of AI agents and human-AI collaboration. Section 3 depicts the research gaps and questions that guide my dissertation. In Section 4, I elaborate on the research design and methodological approach employed in this dissertation. In Section 5, I summarize the main results and of each of the six essays included in this dissertation. Lastly, Section 6 concludes the dissertation by discussing the key findings, addressing the limitations of the study, and highlighting opportunities for future research.

In presenting and discussing the individual essays of my dissertation, I use "we" as all essays were co-authored. The following sections include content drawn from these essays, and for the sake of readability, I have omitted standard citation labels.

2 Background

This section establishes the foundation for my essays and introduces the relevant research concepts addressed in this dissertation. In the following, I will present relevant AI concepts, define related terms, clarify AI's potential and shortcomings, and elaborate on how AI alters our understanding of information systems. After that, I will outline the key concepts of human-AI collaboration, including the underlying rationale, potential conflicts, and strategies for ensuring beneficial and efficient collaboration.

2.1 Conceptual foundations of artificial intelligence

Artificial Intelligence (AI) has evolved from a theoretical niche topic to a ubiquitous concept driving innovation across various domains, transforming decision-making processes and operational efficiency (Agrawal et al., 2019; Benbya et al., 2020). Various disciplines, including computer science, social science, philosophy, and others, have approached AI, leading to many different conceptualizations, thus requiring clarification (Russell & Norvig, 2022). What most definitions have in common is their reference to AI as the approach of making machines intelligent and behaving with reference to human capabilities and skills (Brynjolfsson & Mitchell, 2017; Russell & Norvig, 2022). This common ground, however, raises the question of whether a calculator may already be considered artificially intelligent. While most people would have considered the calculator artificially intelligent decades ago, today's prevailing opinion is different. This phenomenon indicates that the concept of AI inherently carries some sort of openness and dynamism (McCorduck & Cfe, 2004). Accordingly, what is perceived as artificially intelligent today may not necessarily be considered as such in the future. Considering this evolving nature of AI, I follow Berente et al. (2021, p. 1435), defining AI as the "*frontier of computational advancements that references human intelligence in addressing ever more complex decision-making problems*". In that view, AI does not refer to a specific technology or a fixed set of capabilities. Instead, AI applications are becoming increasingly intelligent, autonomous, and inscrutable, allowing for an increased scope of use cases (Berente et al., 2021).

In recent years, the technological advancements around machine learning (ML) have fueled the expansion of AI's performance and scope. ML refers to a set of methods enabling machines to solve problems by learning from data instead of being explicitly programmed to do so (Jordan & Mitchell, 2015; Kühl et al., 2022). This paradigm shift in computational information processing has significantly improved AI's ability to perform tasks once considered exclusive to humans. As a consequence, AI-based information systems are becoming increasingly capable of autonomously performing tasks that are complex, dynamic, and uncertain (Baird & Maruping, 2021; Berente et al., 2021). They also become less transparent with greater autonomy and intelligence, complicating our understanding of their internal mechanisms and decision-making processes (Berente et al., 2021). All these advances interrogate our understanding of information systems (IS), prompting us to reconsider the relationship between humans and information technology.

From an IS research perspective, IS artifacts have generally been viewed as subordinate agents that support humans and act on their behalf (Demetis & Lee, 2018; Orlikowski & Iacono, 2001). In that regard, humans used IS artifacts as tools, disacknowledging their agentic capabilities. However, by increasing autonomy and intelligence, IS artifacts' previously subordinate role now gets equal footing as humans and beyond (Dattathrani & De', 2023; Dung, 2024). AI-based IS artifacts become capable of transferring both rights and responsibilities from and to human agents, making them agentic (Baird & Maruping, 2021). Therefore, agentic IS artifacts can now exhibit higher levels of decision-making latitude, particularly in situations that are considered dynamic and uncertain (Baird & Maruping, 2021; Schuetz & Venkatesh, 2020).

AI's increasing capabilities drive its persuasion of business and society. AI enables task automation across many activities and processes, allowing for unprecedented productivity and growth in efficiency (Davenport & Kalakota, 2019). However, complete automation of activities and processes is not always desirable due to ethical, legal, social, economic, or technical constraints (Brynjolfsson et al., 2018; Dellermann et al., 2019; Fügener et al., 2022). For instance, considering today's healthcare processes, even though AI can surpass humans in many diagnostic tasks, the de facto right and responsibility to make medical diagnoses still remains with human doctors due to societal obligations (Brynjolfsson & Mitchell, 2017; Göndöcs & Dörfler, 2024; Martinho et al., 2021). Moreover, many processes include nonmaterial and material components, requiring physical actuation, which most AI systems cannot perform in contrast to humans. Accordingly, rather than striving for full automation, AI is more likely to augment humans' activities and processes, forming a collaborative relationship (Baird & Maruping, 2021; Dellermann et al., 2019; Hemmer et al., 2023).

2.2 Conceptual foundations of human-AI collaboration

Both humans and AI agents inherently possess distinct capabilities complementing the attributes being determined by external constraints such as societal norms or ethical standards (Baird & Maruping, 2021; Dellermann et al., 2019). For instance, humans possess superior intuitive and emotional intelligence and contextual understanding, while AI possesses higher analytical capabilities (Huang & Rust, 2018; Jarrahi, 2018; Wirtz et al., 2018). Moreover, humans and AI agents possess differing sensors and actuators, leading to incongruous interactions with their environment (Russell & Norvig, 2022). Consequently, since humans and AI agents have unique strengths and weaknesses, it becomes desirable that they work together so that their collaborative action can exceed their individual performance (Dellermann et al., 2019; Fügener et al., 2022).

Contemporary developments of AI applications increasingly emphasize collaboration designs that deliberately consider complementary interactions, given their superior capabilities. From a research perspective, such synergetic interaction is considered as *human-AI collaboration*. Following Hemmer et al. (2024) and Vössing et al. (2022), we refer to human-AI collaboration as the *"process in which two or more agents work together to achieve shared goals [...] involving at least one human and at least one computational agent*" (Terveen, 1995, p. 67). Human-AI collaboration can be instantiated in various modes of interaction and is differentiated by a wide range of dimensions, such as, for instance, the agents' hierarchy and involvement, the timing and frequency of interaction, as well as the direction and content of information exchange (Adam et al., 2024; Baird & Maruping, 2021; Hemmer et al., 2021; Hemmer et al., 2023). Currently, the prevailing form of human-AI collaboration is instantiated by designing AI agents as decision-support systems (DSS) that provide the hierarchical superior human with relevant information to facilitate problem-solving and decision-making (Lai et al.; Phillips-Wren, 2013; Power, 2002). In that sense, humans delegate tasks to AI agents by transferring rights and responsibilities for task execution (Baird & Maruping, 2021; Fügener et al., 2022). For instance, in healthcare, doctors provide patient information to an AI-based DSS and ask it to suggest potential diagnoses, which ultimately enhances the accuracy of clinical decisions and improves patient outcomes (Nserat et al., 2023).

The concept of delegation is becoming increasingly relevant in human-AI collaboration, particularly as AI's facets of intelligence and autonomy expand (Baird & Maruping, 2021; Berente et al., 2021). As AI agents' decision-making latitude rises, they can collaborate with humans on equal footing and even acquire hierarchical superiority (Baird & Maruping, 2021).

From a performance perspective, AI-led delegation offers a wide range of opportunities for human-AI collaboration. According to Fügener et al. (2022), humans struggle to accurately assess their own abilities and the difficulty of tasks, resulting in poor decisions in human-AI collaboration. In contrast, the performance of AI artifacts autonomously appraising whether to delegate a task to a human or not tends to be higher, deeming this type of human-AI collaboration economically attractive (Fügener et al., 2022).

Nonetheless, the design of effective human-AI collaboration fostering complementarity is no panacea and represents a subject of extensive scholarly research (e.g., Amershi, Inkpen, et al., 2019; Hemmer et al., 2021; T. Li et al., 2023). Besides ensuring that AI agents are fair (Mehrabi et al., 2022), transparent (Adadi & Berrada, 2018), and robust (Shneiderman, 2020), it is vital that their mutual information exchange is appropriately established (T. Li et al., 2023). One major obstacle in the information exchange between humans and AI agents is humans' susceptibility to under and over-rely on AI agents' advice (Schemmer et al., 2023). For instance, when humans receive advice from an AI agent, they are easily influenced—intentionally or unintentionally—and may exhibit biases resulting in sub-optimal complementarity (Schemmer et al., 2023). Therefore, AI designers should design AI agents enabling appropriate reliance, which I consider as "*the pattern of reliance behavior(s) that is most likely to result in the best human-automation team performance*" (Talone, 2019, p. 13). In doing so, AI agents should provide interpretable explanations of their reasoning, allowing users to develop a calibrated trust in the AI's output (Inkpen et al., 2023), ultimately reducing the uncertainty associated with AI advice (Jiang et al., 2022; Schoeffer et al., 2022; Vössing et al., 2022).

Against this backdrop, the collaboration with increasingly autonomous AI agents further complicates the design of human-AI collaboration. For instance, information asymmetries can intensify due to their differing natures and capabilities (Baird & Maruping, 2021; Vincent, 2021). Additionally, as humans cede control over processes and decision-making to AI, conflicts of interest can arise (Fügener et al., 2022; Vössing et al., 2022).

3 Derivation of research gaps and research questions

To examine the design and management of effective and beneficial collaboration between humans and autonomous AI agents, I propose three overarching research goals collectively addressed through the respective essays in this dissertation. Each essay tackles a unique research question, filling a specific gap and contributing to one of the three overarching goals. Accordingly, I apply thoughtfully chosen research methodologies suited explicitly to each research endeavor. For well-understood yet undertheorized topics such as the *design and management of AI applications* and the *interaction design of human-AI collaboration*, I employ prescriptive approaches. Meanwhile, I adopt more explorative approaches for emerging and scarcely understood phenomena like the *effects of increasing AI autonomy on human-AI collaboration*. This section will illuminate the concrete research gaps and derive the resulting research goals of this dissertation.

3.1 Understanding the design and management of AI applications

AI applications are becoming increasingly autonomous as their learning capabilities improve (Berente et al., 2021). Simultaneously, these applications are growing more opaque, becoming intelligible only to selected stakeholders (Berente et al., 2021). Moreover, drifts in the abstract AI model or the real-world environment can alter the behavior of AI applications over time, posing a risk for inefficient and potentially harmful operations. Accordingly, dedicated coordination and control through comprehensive AI management are vital for successfully operating AI applications (Benbya et al., 2020; Faraj et al., 2018; Jöhnk et al., 2021). Therefore, decision-makers must ensure having all relevant information at hand to make informed decisions, given that information is the "*knowledge for the purpose of taking effective action*" (Mason & Mitroff, 1973, p. 475)

However, facilitating this information exchange is a complex endeavor. Operating AI technologies involves multiple parties, including AI developers, primary users, IT architects, process owners, and legal counsels. Each of these stakeholders brings their unique expertise, goals, and information needs, thereby complicating the design of effective information processing. Particularly, the insufficient satisfaction of information needs leads to information gaps and high levels of task uncertainty (Galbraith, 1974; Haußmann et al., 2012). Accordingly, enhancing information processing capabilities and bridging information gaps among diverse stakeholders are critical tasks for AI management to mitigate the anticipated task uncertainty.

AI managers must "*communicate, lead, coordinate, and control organizational efforts to […] realize their goals, while at the same time, avoiding the negative consequences*" (Berente et al., 2021, p. 1434). Despite its importance, practitioners and researchers face challenges in managing AI applications in their production environment and fail to develop adequate, holistic managerial practices promoting information processing (Ananny & Crawford, 2018; Berente et al., 2021; Dwivedi et al., 2019).

Although significant research on AI application management has emerged (e.g., J. Li et al., 2021; Sturm et al., 2021; Teodorescu et al., 2021), existing work has predominantly focused on conceptualizing individual management factors (i.e., describing what to manage) rather than providing a holistic understanding and exploring the interrelations among these factors. Existing AI management approaches, such as MLOps, are promising, yet they focus on technical facets instead of holistic management (Kreuzberger et al., 2023). Following the calls for research on the management of AI applications in organizations (Baier et al., 2019; Berente et al., 2021; Collins et al., 2021), we ask:

What are the factors of managing AI applications in healthcare and how are they related?[1] *What management practices improve information processing among stakeholders in AI management? (Essay 1)*

Besides improving AI management by facilitating information exchange among stakeholders, another AI management approach is the direct retrieval of management-relevant information from the AI's technology stack, primarily relying on machine learning (ML) technology. Such ML monitoring comprises the comprehensive observation, performance measurement, and analysis of an ML application's behavior in its production environment. Moreover, it facilitates taking appropriate actions when deviations from the application's intended status arise (Arpteg et al., 2018; Breck et al., 2017; Klaise et al.; Lins et al., 2021). ML monitoring can assist organizations in gathering management-relevant information about their applications and minimizing risks associated with their operation in production environments (Breck et al., 2017; Köchling et al., 2021)

However, organizations have limited guidance on monitoring their ML applications from a socio-technical perspective. Established software engineering approaches, such as DevOps, fall short of addressing the unique specifics of ML (Amershi, Begel, et al., 2019). While software

[1] In Essay 1, I situate the research question within the healthcare domain owing to its operational complexity, the critical nature of its processes, and the diverse characteristics of its stakeholders. The insights gained allow for application across other domains.

engineering focuses on discrete artifacts with specific code and concepts (van den Heuvel & Tamburri, 2020), ML applications also face data-related and model-related issues on top of known code complexity issues (Amershi, Begel, et al., 2019; Arpteg et al., 2018). Existing ML management approaches deliberately consider ML monitoring, yet they focus on technical monitoring facets within the ML pipeline rather than taking a socio-technical perspective (cf. Kreuzberger et al., 2023). Since ML monitoring is more than a monitoring software system, identifying the necessary practices for monitoring ML applications in production environments remains an unanswered question. Given the broad spectrum of challenges in ML deployment and the lack of comprehensive research in ML application management, our goal is to further enhance the *understanding of managing AI applications*. Therefore, we ask:

What are relevant practices for monitoring ML applications in their production environments? (Essay 2)

In addition to managing the deployed applications, AI managers must ensure that their organizations possess the necessary AI capabilities for operational effectiveness and to stay abreast of future AI innovations (Fischer & Beimborn, 2022; Weber et al., 2023). The AI application management entails a proactive role in governing and enhancing organizational AI capabilities, requiring managerial decisions on which AI capabilities to develop internally versus those to acquire through external sourcing. The rapid technological evolution and the substantial investments needed for in-house development of AI capabilities have led organizations to increasingly favor external sourcing (Wei & Pardo, 2022), such as from so-called *AI service platforms*. Despite the growing reliance on these platforms, a systematic characterization of AI service platforms remains an unaddressed gap in academic research and practical application. It is unclear which attributes define AI service platforms and how AI managers can utilize them to streamline and inform decisions regarding the organizational integration of AI services to acquire relevant AI capabilities. Building upon our investigation into the design and management of AI applications, we pose the following question:

Which essential properties characterize the multitude of AI service platforms in practice? (Essay 3)

3.2 Understanding the interaction design of human-AI collaboration

When humans and AI agents collaborate on a task, their collective performance can surpass their individual performance (Dellermann et al., 2019). To achieve such complementarity, it is vital for humans and AI to effectively exchange information, enabling them to understand each other and capitalize on their individual strengths while compensating for their weaknesses. Nevertheless, establishing reliable and efficient human-AI collaboration is a non-trivial challenge. This is because AI agents' cognitive processes are opaque and differ fundamentally from humans' cognitive processes (Berente et al., 2021). In this context, various socio-technical factors that shape the interaction design between humans and AI agents influence their complementarity (Bansal et al., 2021; Hemmer et al., 2021; Q. Zhang et al., 2022). However, if these factors are not adequately considered, they can hinder this complementarity, leading to biases (Amershi, Inkpen, et al., 2019; Hemmer et al., 2021; Inkpen et al., 2023). These biases, in turn, may cause over-reliance or under-reliance on AI advice (Buçinca et al., 2021; Hemmer et al., 2021; Schemmer et al., 2023). Consequently, such biases can lead to sub-optimal decision outcomes, thereby distorting the effectiveness of human-AI collaboration (Benda et al., 2021; Buçinca et al., 2021; Schemmer et al., 2023).

A prevalent bias in human-AI interactions is the automation bias, which may emerge when an AI agent offers advice before a human has made a preliminary decision (Cummings, 2017; Schemmer et al., 2022). Considering the provided AI advice before forming an independent evaluation may reduce cognitive dissonance by seeking confirmatory evidence that aligns with the AI's suggestion (Cummings, 2017; Schemmer et al., 2022). Consequently, humans might not evaluate all available information appropriately. Instead, they give undue weight to information that aligns with the AI agent's advice (Parasuraman & Riley, 1997; Sujan et al., 2019). To mitigate automation bias, the designer of the hybrid system may ensure that AI advice is not disclosed until the human has made an evaluation. Instead, AI advice should be presented after the human has made the initial decision. However, integrating the AI advice solely as a control instance may induce contrary effects, such as biases arising from escalation of commitment and algorithmic aversion. Escalation of commitment describes humans' tendency to justify their initial assessment over another conflicting outcome (Staw, 1981). In human-AI collaboration, this phenomenon is closely related to algorithmic aversion, which describes a human preferring human advice over algorithmic advice, although the algorithm is superior (Jussupow et al., 2020). The reason is humans' biased appraisal of the AI's capabilities and outcome (Inkpen et al., 2023; Jussupow et al., 2020). As a result, people tend to undervalue AI advice, ultimately leading to under-reliance on it (Inkpen et al., 2023).

Besides the timing of AI advice, also the supplementary information content (i.e., degree of explanation) of an AI advice can affect the human-AI collaboration (Buçinca et al., 2021).

Assuming that the AI advice comprises an explanation, it can be better evaluated and is more knowledgeable (Bansal et al., 2021), ultimately increasing human reliance on the AI suggestion (Jacobs et al., 2021; Y. Zhang et al., 2020). However, it can also lead to adverse effects. For instance, when the AI advice is false yet provides a convincing explanation, it may induce overreliance, impeding optimal human-AI complementarity.

Achieving human's appropriate reliance on AI advice requires the designer of the human-AI system to balance over- and under-reliance (Benda et al., 2021; Schemmer et al., 2023; Schoeffer et al., 2022). Therefore, it is essential to deliver AI advice at the right time (J. Yin et al., 2020) and include the appropriate explanatory content (Benda et al., 2021; Buçinca et al., 2021; Schemmer et al., 2023). While several scholars have called for research on the effects of specific human-AI interaction designs, there have been only a few quantitative investigations on the effects of AI advice timing and explanatory information on human-AI collaboration (Amershi, Inkpen, et al., 2019; Gerlach & Kuo, 1991; J. Li et al., 2021). To address the uncertainty about the effects of relevant interaction design dimensions, our research goal is to *understand how the timing and explanatory information of AI advice provisioning affect the human-AI collaboration (cf. Essay 4).*

3.3 Understanding the effects of increasing AI autonomy on human-AI collaboration

Recent technological advancements around AI have significantly expanded its performance scope, leading to increasingly autonomous AI agents and novel forms of human-AI collaboration (Peeters et al., 2021). Autonomous AI agents possess superior intelligence, enabling them to perform tasks that are complex, dynamic, and uncertain (Berente et al., 2021; Jarrahi, 2018). Existing IS research has recognized AI agents primarily as subordinate IS agents that support human agents and act on their behalf (Baird & Maruping, 2021; Orlikowski & Iacono, 2001). However, owing to their increasing autonomy, AI agents are no longer limited to performing tasks on humans' behalf (Baird & Maruping, 2021; Dattathrani & De', 2023). Instead, they acquire task ownership and autonomously guide human actions (Baird & Maruping, 2021; Harms & Han, 2019; Wesche & Sonderegger, 2019). The autonomy gain induces a reversal in the direction of task delegation between humans and AI agents, enabling AI agents to delegate tasks to humans.

This new phenomenon of AI agents autonomously delegating to humans poses significant implications for human-AI collaboration. On the one hand, research suggests that human-AI systems can achieve better results when AI artifacts delegate tasks to a human rather than vice versa (Fügener et al., 2022). The reason is that humans tend to be less adept at evaluating their own and the AI agent's capabilities in relation to the difficulty of specific tasks and the expected outcomes of delegation (Fügener et al., 2022). In contrast, AI artifacts possess a superior ability to appraise the capabilities and expected outcomes of delegation (Fügener et al., 2022). On the other hand, AI-led collaboration can foster uncertainty, transforming existing human-AI collaboration structures. This is particularly relevant in organizational contexts, where fundamental shifts in process design and organizational coordination may occur (Benbya et al., 2020; Wesche & Sonderegger, 2019). The uncertainty manifests, for instance, in unclear oversight and accountability, leading to a loss of control over the process outcome for humans when transferring the delegation ownership to autonomous AI agents (Fügener et al., 2022; Steffel et al., 2016). Moreover, the opacity of AI's behavior and underlying reasoning can result in a lack of trust and human aversion against the autonomous AI agent (Vössing et al., 2022). Similarly, the willingness of users to adopt a technology decreases with their perceived increase in autonomy (Chao et al., 2016; Leyer & Schneider, 2019; Złotowski et al., 2017).

While research has extensively examined the human-AI collaboration with humans in the lead, the consequence of AI primacy is vastly unexplored. Existing research lacks comprehensive theorizing of the effects of autonomous AI agents that delegate to human agents in collaborative environments, nor does it capture the implications for organizations. We argue for dedicated research on the emerging phenomena of autonomous AI agents, as recent technological advancements are already enabling autonomous AI agents, and future developments will further fuel their adoption (cf. Dellermann et al., 2019; Hemmer et al., 2021; Vössing et al., 2022; Xu et al., 2023). Aligned with the overarching goal of my dissertation, which is to navigate the collaboration between humans and autonomous AI agents, I aim to enhance the *understanding of the effects of increasing AI autonomy on human-AI collaboration*. I approached the research goal with two distinct research endeavors (i.e., Essay 5 and Essay 6). Essay 5 sheds light on how human-AI collaboration evolves when autonomous AI agents take over delegation ownership, reducing the uncertainties regarding such collaboration faced by companies. Therefore, I ask: *What are the key tensions that arise from delegating tasks from artificial intelligence to humans, and what factors contribute to these tensions?(Essay 5)*

In contrast, Essay 6 focuses on dyadic human-human relationships that are evolving by integrating autonomous AI agents. We propose that the dyadic human-human relationship is transforming into a triadic relationship with autonomous AI agents. We ask: *How do agentic IS artifacts affect the dyadic interaction relationship of humans? (Essay 6)*

4 Dissertation structure and research designs

This dissertation comprises six essays complemented by the introduction section. Based on a cumulative approach, each essay comprises a completed research activity with a dedicated research question, empirical data, research method, and outcome that collectively contribute to the aforementioned research goals. The research essays have been published or are currently under review in renowned journals and conferences in information systems research. All publication outlets adhere to the highest standards of good research practices, including a rigorous double-blind peer review process that ensures the integrity and quality of scholarly work. In Table 1, I provide an overview of the essays, their publication outlet, and their current status. Furthermore, I provide the individual rankings of the respective journals and conferences based on the latest rating of the German Academic Association for Business Research (VHB).

Table 1. *Overview of the essays in this dissertation*

Essay	Title	Publication Outlet and Status	VHB-2024
1	Managing artificial intelligence applications in healthcare: Promoting information processing among stakeholders	International Journal of Information Management (IJIM) *Status: Published*	B
2	What gets measured gets improved: Monitoring machine learning applications in their production environments	IEEE Access *Status: Published*	B
3	Gateways to artificial intelligence: Developing a taxonomy for AI service platforms	Proceedings of the European Conference on Information Systems (2021) *Status: Published*	A
4	Improving decision accuracy in human-AI collaboration: The role of timing and explanatory information	Information Systems Journal *Status: Under review*	A
5	Task delegation from AI to humans: A principal-agent perspective	Proceedings of the International Conference on Information Systems (2023) *Status: Published*	A
6	Toward triadic delegation: How agentic IS artifacts affect the patient-doctor relationship in healthcare	Journal of the Association for Information Systems *Status: Conditionally Accepted*	A

Information systems research is an interdisciplinary science examining social contexts and information technology (Grover & Lyytinen, 2015; Thatcher et al., 2018). Understanding how information technology shapes social constructs and vice versa is at the core of IS research, delineating it from the relatively naturalistic computer science and social sciences (Lee, 2001; Thatcher et al., 2018). The rapid persuasion of information technology within society has provided momentum to the information systems discipline over the past decades (Hirschheim & Klein, 2012). However, owing to the complexity of socio-technical phenomena, information systems research has been struggling with creating grand theories (Lyytinen & King, 2004). As a result, IS research has mostly focused on applying existing grand theories from other disciplines to emerging IS phenomena, resulting in mid-range theory (Gregor, 2006; Grover & Lyytinen, 2015). The application of such a diverse set of theories, along with a wide range of qualitative and quantitative methodologies and paradigms, has laid the foundation for Information Systems (IS) research as a pluralistic research discipline (Mingers, 2001). Throughout my research, I adopt the spirit of pluralistic research and apply different research methods and paradigms, allowing for more affluent and more reliable research outcomes. In doing so, I combine positivist and interpretivist research approaches to achieve my research goals (Goldkuhl, 2012). In the following, I provide an overview of the research designs of the essays in this dissertation.

In Essay 1, we examined the management of AI applications from an organizational perspective. We contextualized our research within the healthcare domain due to its operational complexity and the critical nature of its services, alongside the diverse priorities of its stakeholders. We followed a qualitative two-step research approach consisting of a structured literature review and a subsequent interview study through the theoretical lens of the organizational information processing theory (OIPT) (Galbraith, 1974; Haußmann et al., 2012). Combining structured literature reviews and interview studies is common in qualitative research studies (e.g., Baier et al., 2019; Benner et al., 2022; Gimpel et al., 2018). The body of literature captures existing theoretical knowledge, which is then extended, validated, and triangulated through expert knowledge and experience (i.e., interview study). Considering the OIPT enables us to examine the management of AI applications from an information processing perspective. We adopt this approach because managing AI applications is a multi-stakeholder process that constantly requires interaction among multiple specialized inter- and intra-organizational teams. For the literature synthesis, we followed a concept-centric approach based on Webster and Watson (2002) and applied rigorous literature coding based on Gioia et al. (2013). We aimed to identify, analyze, and structure AI management factors representing abstractions of core AI management tasks and managerial interrelations. Moreover, we derived practices facilitating information processing among the stakeholders, levering AI application management. Based on our literature-based findings, we then iteratively developed the AI application management (AIAMA) model that theorizes the management factors and managerial interrelations. The goal of synthesizing the AIAMA model was to develop an "underlying structure, the scaffolding or frame" (Merriam & Tisdell, 2016, p. 85) for AI management theory. After that, we conducted eleven expert interviews to further refine the AIAMA model and enhance underlying management practices with expert knowledge from research and practice. We used the criteria proposed by March and Smith (1995, p. 261) to validate the model abstraction during the interviews. We asked the experts for feedback on the model's "*fidelity with real-world phenomena, completeness, level of detail, robustness, and internal consistency*". After completing the interview study and achieving a mature state of the AIAMA model, we applied it to gather further insights into AI management practices from an organizational perspective. In doing so, we analyzed the information processing across the various AI management functions. For this purpose, we utilized the AIAMA model to analyze the information processing based on 34 exemplary management tasks related to the respective management factors. The modeling of these factors within the AIAMA model was carried out during a research workshop within the research team, aiming to ensure scientific objectivity. Applying the AIAMA model enabled us to validate its applicability and elaborate on three focal patterns concerning information processing during AI management practices. Moreover, applying the model enhanced the comprehensibility of the AIAMA model itself, by demonstrating how the AIAMA model relates to the AI management problem and how it can be used to explain the underlying mechanisms.

In Essay 2, our goal was to derive ML monitoring practices and offer a comprehensive overview of the necessary steps for successful ML application monitoring endeavors. Adopting a rigorous qualitative research methodology as outlined by Bhattacherjee (2012) we grounded our work in the principles of intelligent agent theory (Rudowsky, 2004; Russell & Norvig, 2022; Wooldridge & Jennings, 1995). Given the nascent research stage of ML monitoring, we approached our research question exploratively. Therefore, we conducted an interview study with ML practitioners to gain insights into monitoring practices for ML applications. Following Myers and Newman (2007) and Schultze and Avital (2011), we conducted semi-structured interviews. The expertise of the practitioners captured through the interviews allowed us to better understand the challenges of monitoring productive ML applications and identify current best practices. We analyzed the transcribed interviews using the coding approach proposed by Saldaña (2021). Additionally, we conducted a multivocal literature review on ML monitoring, including both academic and grey literature, as proposed by Ogawa and Malen (1991) and

Garousi et al. (2019). The multivocal literature allowed us to review and expand the identified ML monitoring practices through our interview study. Lastly, we collected data on existing ML monitoring tools, complementing our findings with further practical insights into real-world monitoring practices. Based on our comprehensive data analysis, we derived 17 monitoring practices to successfully monitor organizations' ML applications. We organized the practices according to the five steps of a typical quality management cycle to sequentially order the practices for improved operationalization. The cycle consists of five steps: define, measure, assess, act, and control (Mast & Lokkerbol, 2012; Tonini et al., 2006). The constructed monitoring framework fosters the effective operation of ML applications in their production environments, hence contributing to the academic discourse on AI application management from a technology perspective (Berente et al., 2021; Hummer et al., 2019; Kreuzberger et al., 2023)

Essay 3 complements my research on AI management by shedding light on AI platforms' design dimensions and properties. To structure and analyze the concepts of AI service platforms, we developed a taxonomy following the guidelines of Nickerson et al. (2013). A taxonomy represents a form of classification in which one derives a system of construct groupings, either conceptually or empirically (Nickerson et al., 2013). In IS discipline, taxonomies are an established and reliable approach to structure emerging research phenomena (e.g., Jöhnk et al., 2017; Lösser et al., 2019; Püschel et al., 2016). Accordingly, we found it appropriate to develop a taxonomy to better understand the still-elusive phenomenon of AI service platforms. The iterative method proposed by Nickerson et al. (2013) recommends starting with defining an overarching meta-characteristic that guides the creation of the taxonomy's characteristics and dimensions. For our meta-characteristic, we considered scholars and practitioners as our target group, aiming to understand the various concepts of platform-based AI services to make informed decisions. Therefore, our meta-characteristic was "*essential properties characterizing of AI service platforms in practice*". Furthermore, we determined the ending conditions for taxonomy development, ensuring the formal correctness of the taxonomy and its usability. We defined the ending conditions as follows: First, the analyzed sample of constructs had to be relevant. Second, objects were neither split nor merged during the last iteration, and no new characteristics or dimensions were added, split, or merged. Third, every dimension, characteristic, and cell had to be free of redundancies. After that, the iterative taxonomy development process commenced. Each iteration followed either a conceptual-to-empirical or an empirical-to-conceptual approach. The conceptual-to-empirical approach relies on existing literature and the researcher’s knowledge and experience, whereas the empirical-to-conceptual approach involves analyzing a sample of available objects to extract shared characteristics.

For our first iteration, we began with an empirical-to-conceptual approach, conducting an online search to sample AI service platforms available on the market. We identified a total set of 31 AI service platforms offered by start-ups and incumbent organizations. For the initial iteration of taxonomy development, we selected a random sample of 15 AI service platforms. To refine the initial state of our taxonomy, we adopted a conceptual-to-empirical approach in the subsequent iteration. In doing so, we contextualized existing conceptual knowledge by drawing from the relevant literature. Consequently, we expanded, modified, and merged various dimensions and features within our taxonomy and conducted a second empirical-to-conceptual iteration employing an extended sample of AI service platforms.

After an unsatisfactory evaluation of the ending conditions, we conducted semi-structured expert interviews to augment our primary data (Jöhnk et al., 2017; Myers & Newman, 2007). These interviews significantly contributed to the development of the taxonomy by highlighting certain shortcomings in specific dimensions (e.g., ambiguous wording or missing characteristics) and suggesting opportunities for improvement (e.g., enhancing the presentation of the taxonomy). Leveraging this new primary data enabled us to address the maturity issues within our taxonomy. To ensure our taxonomy aligned with existing AI service platforms, we undertook another empirical-to-conceptual iteration, which facilitated the classification of our complete sample of AI service platforms. Following minor adjustments in the subsequent iteration, we sought and integrated feedback from scholars and taxonomy experts. Subsequently, another conceptual-to-empirical iteration allowed us to validate our taxonomy considering the revisions made. In the final iteration, no further changes to the taxonomy were necessary, as it met all the ending conditions.

Moving towards the second research goal of my dissertation, Essay 4 enhances the theoretical understanding of human-AI interaction and proposes design principles for optimal AI advice provisioning promoting human-AI complementarity. This study explores how relevant design modes for AI advice provisioning impact the overall decision accuracy in human-AI systems. Therefore, we conducted an experiment that allowed us to quantify the causal effects of specific design variables of AI advice provisioning. Experiment research design has demonstrated the viability of expanding theoretical understanding of human-AI interactions across various seminar articles (e.g., Bondi et al., 2022; Fügener et al., 2021; Reverberi et al., 2022).

As a starting point, we set out to formulate research hypotheses, which are clearly defined problem statements allowing for direct verification through empirical investigation (Lazar et al., 2017). Relying on previous scholarly work in human-computer interaction (HCI) research (e.g., Buçinca et al., 2021; Gregor & Benbasat, 1999; J. Yin et al., 2020), we hypothesized that both the timing of AI advice as well as the explanatory information of AI advice affect the decision accuracy of the human-AI system. To empirically test our hypotheses, we conducted an online experiment simulating the decision process of a human being supported by an AI-enabled decision support system (DSS). In choosing an appropriate experimental design to measure relevant effects, we proposed three criteria for selecting a suitable decision-making case: First, the decision-making process should require expert knowledge. Second, the decision tasks should include challenging tasks (i.e., edge cases) to allow the AI-enabled DSS to make a meaningful contribution to human decision-making. Third, the simulated decision process should replicate a time-sensitive decision task, ensuring that humans apply varying cognitive processes (i.e., type 1 and type 2 thinking). Based on the three criteria, we identified football referees' assessment of game situations as a promising decision process for our experiment. Football referees are experts at assessing game situations and making time-sensitive decisions based on situational information combined with their profound experience (Gottschalk et al., 2022). Additionally, football referees frequently assess complex situations that are considered edge cases.

The experiment comprised an introduction, a decision-making experiment, and a post-survey. In the introduction, we informed the participant about the nature of the experiment and asked for the participant's consent concerning the experiment's procedures and data processing. After that, we sequentially presented football game situations to the participants, asking them to review tacklings potentially denying a goal or an obvious goal-scoring opportunity (DOGSO situations) in cooperation with an AI-based DSS. Assessing DOGSO situations requires referees to weigh more than 40 decision parameters for decision-making (e.g., location of the foul, control of the ball, distance between player and goal, etc.), underpinning its potential for decision support (Gottschalk et al., 2022; The International Football Association Board, 2023). The decision-making experiment consisted of four setups, each comprising five game situations. The four setups differed in the timing of the AI advice and the presence of explanatory information (i.e., four treatments in a 2x2 design, within-participant design). The possible classification of each case was either *no foul*, *foul and no card*, *foul and yellow card,* or *foul and red card*. Finally, to capture control variables, the participants answered a post-survey, querying the participants' attitudes to AI in general (AIG) and to AI in sports refereeing (AIR) as well as socio-demographics and experience in the respected medical area.

We tested our hypotheses using regression analysis with a total of seven different regression models. Models 1 to 4 were straightforward tests for hypotheses (i.e., H1 to H4). The regression models 5 to 7 further added to this by analyzing the role of cognitive effort and trust as mediators. Before conducting our analysis, we performed a confirmatory factor analysis (CFA) to validate the robustness of the reflective measurement models for trust, effort, general attitude toward AI, and attitude toward AI in sports refereeing

In Essay 5 and Essay 6, I aimed to expand the understanding of the effects of increasing AI autonomy on human-AI collaboration. In Essay 5, I focused on exploring how human-AI collaboration evolves when AI takes over the ownership of task delegation. We studied the paradigm shift of AI artifacts becoming the delegating agent through principal-agent theory (PAT) as our theoretical lens. The PAT was initially developed to describe relationships as contracts where "*one or more persons (the principal(s)) engage another person (the agent) to perform some service on their behalf which involves delegating some decision making authority to the agent*" (Jensen & Meckling, 1976, p. 5). We argue that the principal-agent theory suits our study purpose well because it elucidates the nature of cooperation and conflict between humans and AI artifacts as agentic entities, with both collaborating to achieve objectives (Baird & Maruping, 2021; Emirbayer & Mische, 1998). In this context, related research has already applied PAT within the human-AI realm, albeit with a focus on the relationship between humans as principals and algorithms serving as agents (e.g., Borch, 2022; Kim, 2020)

Our methodological approach was twofold: We combined a structured literature review guided by Webster and Watson (2002) with qualitative semi-structured interviews Myers and Newman (2007) to integrate insights from theory and practice and to address our research objective. In doing so, we first synthesized existing literature to get an overview of the current state of research on delegating relationships between humans and AI artifacts. While the research has predominantly focused on human-to-AI delegation, considering AI only as the human principal's agent, we were able to capture existing knowledge on delegation structures, mechanisms, and factors that affect and determine the delegation design between humans and AI artifacts from the relevant papers in the structured literature review.

We then built upon the identified state of the research by conducting an in-depth interview study to enhance our theoretical understanding of the phenomenon of AI artifacts being in the principal role and delegating ownership. For the interview study, we considered two types of experts. First, we aimed to interview individuals acquainted with autonomous systems in practical settings, preferably those engaged in implementing agent systems with AI artifacts possessing high degrees of autonomy and task responsibility. Secondly, as we aim to explore the manifestation of the principal-agent relationship between humans and AI when AI assumes the principal role, we also sought experts in principal-agent theory, particularly individuals with experience dealing with digital, nonhuman agents. In total, we conducted 13 semi-structured interviews.

For data analysis, we adhered to the systematic approach proposed by Gioia et al. (2013), which is frequently employed when a more profound comprehension of organizational processes and dynamics is necessary. This method is well suited to our research objective. It allows unstructured qualitative datasets to be processed, relevant categories and relationships between them to be formed, and new concepts, ideas, and theories to emerge (Corbin & Strauss, 2015; Gioia et al., 2013). We adopted the iterative three-step coding process (i.e., open, axial, and selective coding) to study the phenomenon of interest at different abstraction levels (Gioia et al., 2013). For structuring the developed codes, we adhered to the concepts of the principal-agent theory guiding the development of the aggregate dimensions, as our primary objective was to investigate the PAR in AI-to-human delegation.

Lastly, Essay 6 investigates how agentic IS artifacts affect dyadic human interaction relationships by evolving into a delegation triad. We opted for phenomenon-based theorizing, which combines deductive and inductive theorizing (Fisher et al., 2021; Gregory & Henfridsson, 2021). Phenomenon-based theorizing focuses on studying emerging phenomena that are difficult to understand with existing theory or that conflict with such (Fisher et al., 2021). For instance, phenomena induced by the transformative influence of information technology—such as AI—are particularly suitable for phenomenon-based theorizing (Gregory & Henfridsson, 2021; Krogh, 2018). In our case, the evolution from dyadic human delegation relationships to a triadic relationship through an agentic IS artifact marks a new phenomenon that changes agentic relationships and behaviors.

We study our phenomenon within an exploratory single case of an agentic IS artifact from a health technology company that addresses patients with incontinence, specifically those with neurogenic bladder dysfunction. We investigate the interactions between patients, doctors, and an agentic IS artifact through the theoretical lens of delegation, which is also the core research stream we seek to expand through our phenomenon-based theorizing. We rely on the delegation framework of Baird and Maruping (2021), examining how an agentic IS artifact affects the patient-doctor relationship and theorizing novel agentic behaviors, including role behaviors, interaction patterns, and social constructs.

For the data collection, we employed a longitudinal, multi-source approach primarily consisting of technical documentation, corporate documentation on stakeholder interaction, and interviews with patients, doctors, and AI delegation experts (Dubé & Paré, 2003; Walsham, 1995). We gathered 50 recorded and transcribed interviews and notes from more than 100 face-to-face meetings, phone calls, emails, and instant message conversations with patients, doctors, and delegation experts. We also analyzed over 100 pages of technical documentation and notes from two observations of a patient-doctor appointment and 14 site visits to a spinal cord center, a healthcare center, and a urology center.

For our data analysis, we followed the guidelines of Gioia et al. (2013). We analyzed the interview data in three successive coding stages (i.e., open coding, axial coding, and selective coding). The first coding stage involved extracting codes directly from our informants' spoken words (i.e., patients, doctors, and delegation experts), using minimal interpretation (Gioia et al., 2013). During stage 2 of our analysis, we examined the preliminary results and began to identify emerging themes. We reviewed the codes and interview transcripts iteratively and grouped the data into broader themes that connected several concepts. After consolidating the second-order themes, we further abstracted them into aggregated dimensions. In coding stage 3, we built and refined our aggregated dimensions, relating them to our theoretical lens in multiple discussion rounds.

As the final step, we revisited existing theories, critically appraised them and engaged in a dedicated activity to further theory development. This approach ensured that our phenomenon under study was effectively captured by theory.

5 Summary of results

I will now summarize the results of the essays, guiding the management and design of human-AI collaboration in organizational contexts and in light of increasing AI autonomy.

5.1 Essay 1: Managing artificial intelligence applications in healthcare: Promoting information processing among stakeholders

Our comprehensive literature analysis and in-depth interviews identified 32 relevant factors for managing AI applications. The management factors represent theoretical constructs that induce management tasks into the management sphere and require effective information processing for resolution. Every management factor encompasses multiple management tasks that may emerge due to changes in environmental conditions or the AI application. Besides the management factors, we further derive valuable information on the design of management roles, organizational capabilities, and operational practices in AI application management. We structure managerial requirements along three levels, i.e., organization-, role-, and task-level.

Based on our empirical data, we iteratively developed the AIAMA model. The management factors represent *what* to manage, and management practices depict *how* to manage for improving information processing among stakeholders in AI management. Therefore, we rely on the management factors as the foundational building blocks of our model. We consider these factors the primary sources of task uncertainty and equivocality, leading to information processing challenges among stakeholders. The managerial practices are instantiated through five management cycles: technical AI management, contextual management, process management, user requirements management, and integration management. In doing so, we also integrate the concept of information processing through distinct management cycles, enabling coordination and control of the AI management task arising.

By applying the AIAMA model to the 34 management tasks, we observed three focal patterns concerning information processing during AI management practices. Firstly, we recognized that the initial identification of management tasks often occurred in different locations from where they needed to be addressed. Secondly, we could observe that management tasks impacting user requirements, such as transparency issues, often involved two distinct management aspects: the predominantly technical factor of AI application and the healthcare-specific factor of the medical process. Thirdly, we observed a difference between the top-bottom and left-right management functions. We often found challenges and their solutions along the model's horizontal axis related to AI applications and processes. In contrast, the vertical axis, covering user needs and contextual restrictions, tended to define the boundaries within which we managed specific issues.

The essay’s results improve the operation of AI applications by providing a comprehensive framework for AI management, illuminating organizational governance structures and information flows among stakeholders. Accordingly, the essay acknowledges AI management as a system of inter- and intra-organizational interactions characterized by specialized and siloed competencies that create information barriers and task uncertainty.

5.2 Essay 2: What gets measured gets improved: Monitoring machine learning applications in their production environments

Essay 2 proposes a structured approach to monitoring ML applications, combining insights from a qualitative interview study comprising a comprehensive literature review of both academic and grey literature as well as a interview study and ML tool review. The research’s key result is identifying five characteristics of ML production environments and developing 17 specific monitoring practices. These practices are organized within a quality management cycle consisting of define, measure, assess, act, and control phases, each encompassing several targeted actions. This systematic approach aims to provide organizations with a clearer understanding of navigating the complexities of monitoring ML applications, ensuring they remain effective, reliable, and aligned with business objectives over time.

Our findings further indicate that monitoring ML applications requires a comprehensive approach beyond traditional software monitoring techniques. This need arises due to the unique challenges associated with ML applications, such as dealing with data drift, addressing model bias, and managing the impacts of environmental changes on ML performance. Moreover, the research underscores the importance of integrating a combination of technical, operational, and business metrics to comprehensively control the performance and effective operation within their production environment.

The results highlight the dynamic and interconnected nature of ML application monitoring. Given the constant evolution, implementing ML monitoring practices is crucial for organizations’ ability to manage ML applications proactively, ensuring they remain robust and deliver maximum value.

5.3 Essay 3: Gateways to artificial intelligence: Developing a taxonomy for AI service platforms

The primary result of Essay 3 is a comprehensive taxonomy to systematize AI service platforms based on existing literature, expert interviews, and a sample of 31 AI service platforms. The taxonomy is structured into three overarching layers: *customer context*, *AI service platform's offering*, and *third-party integration*, comprising 11 dimensions. The layer of customer context captures the contextual fit between the AI service platform and the customer. AI service platforms' offering comprises functional service dimensions that match the customer's needs with the AI service platform's technological offerings. Third-party integration depicts the access and integration of third parties contributing to the platform. Overall, the taxonomy highlights the modular design of AI service platforms, accommodating various use cases through customizable service configurations.

By systemizing the AI service platforms' characteristics into a taxonomy and applying the taxonomy to existing AI service platforms, we further identified the prevailing motives of AI service platforms. Additionally, we rigorously developed a definition of AI service platforms. In doing so, we identified three prevailing motives behind AI service platforms: enabling organizations to create their own AI models by providing high degrees of freedom, offering ready-to-use AI applications that reduce development effort, and focusing on secondary AI services like complementary services or resources. Moreover, applying the taxonomy to our sample set shows that most AI service platforms are so diverse that they do not fully comply with existing platform definitions. Accordingly, based on the development and application of our taxonomy and building upon existing platform definitions (Gawer, 2009), we could characterize AI service platforms as follows: *AI service platforms provide organizations with access to AI technology to support them in creating or using AI applications through federating and coordinating constitutive agents, leveraging value by enabling economies of scope, and entailing modular technological architecture.*

In conclusion, the paper provides a foundational step toward a common understanding of AI service platforms. It offers valuable insights for future research directions, including the evolution of these platforms, their integration into organizational contexts, and the development of higher-order theories based on the taxonomy. Thus, the developed taxonomy not only aids in understanding the diverse offerings and structures of AI service platforms but also serves as a decision-making tool for practitioners, helping them select suitable platforms based on specific organizational needs and use cases.

5.4 Essay 4: Improving decision accuracy in human-AI collaboration: The role of timing and explanatory information

In Essay 4, we explored how the interaction design from AI-based DSS influences the decision accuracy of human-AI systems. We hypothesized that the timing of AI advice and providing explanatory information alongside AI advice significantly impact decision accuracy. To test the hypotheses, we performed an online experiment with football referees assessing a set of game situations followed by a complementing survey. Each referee assessed five game situations across four interaction modes which differed in the timing of advice (i.e., prior vs. posterior) as well as in the provisioning of explanatory information (i.e., presence vs. absence).

In total, 48 football referees completed our experiment with an average refereeing experience of 11.208 years (SD = 7.294 years). Referee's benchmark decision accuracy was 61.3%, while the accuracy of the DSS was 80%. The collaboration between the referees and the AI-based DSS led to an increase of decision accuracy across all four modes compared to the benchmark decision accuracy. The highest decision accuracy was in mode B2 (i.e., AI advice provided after the initial human assessment, with explanatory information), where referees achieved an average accuracy of 75.8%. Conversely, lowest accuracy was observed in mode B1 (i.e., AI advice after the initial human assessment, without explanatory information), with an average decision accuracy of 69.6%.

Our inferential analysis (i.e., linear regression analysis) revealed a significant positive relationship between explanatory information and decision accuracy for the interaction mode B2, when AI advice is presented after the initial human assessment with explanatory information (.324, $p < .05$). Moreover, the presence of explanatory information was found to significantly increase trust (estimate = .376, $p < .01$), indicating that when referees receive additional explanatory information, their trust in the AI advice increases. When considering both trust and explanatory information, our analysis also reveals that only trust significantly influences decision accuracy (.237, $p < .05$). This indicates that the effect of explanatory information on decision accuracy is mediated by trust.

The experiment contributes to a better understanding of interaction designs for human-AI collaboration concerning the timing and explanatory information of AI advice provisioning.

Second, it increases understanding of the cognitive processes in human decision-making. Practically, our findings support the development of more effective AI-based DSS, promoting better integration of human and AI capabilities for improved decision-making outcomes.

5.5 Essay 5: Task Delegation from AI to Humans: A Principal-Agent Perspective

Essay 5 explores the increasing AI autonomy in human-AI collaboration, focusing on scenarios where AI takes the lead in delegating tasks to humans. Based on a structured literature review and interview study, we investigated the complexities, tensions, and challenges that emerge in this new AI-led delegation through the principal-agent theory as our theoretical lens.

Our results shed light on how different information asymmetries between AI artifacts and humans impact the delegation decision and place specific demands on the delegation relationship. We identified four factors that conceptualize the principal-agent relationship in AI-to-human delegation. We structured our findings along the three existing causes and influences of principal-agent problems, complementing them with a fourth construct specific to AI-to-human delegation. Our results indicate that there are various *information asymmetries* between AI artifacts and humans affecting the delegation decision and placing specific demands on the delegation relationship. Also, we depict *conflicts of interest* arising between AI artifacts and humans. In the third dimension, we present the influences of the *environment and exogenous factors* on the delegation relationship. Complementing the three existing dimensions of principal-agent theory, we introduce a fourth dimension: *human attitude towards the AI artifact as principal*. This construct emerges as a novel factor causing principal-agent problems specifically within AI-to-human delegation contexts. Our results have shown that a human's perception of and attitude toward an AI artifact significantly influences whether or not they follow its instructions. Lastly, we identify several new phenomena across the different dimensions unique to AI-to-human delegation, leading to principal-agent problems in these contexts.

Overall, we propose that the principal-agent problems arising from AI-to-human delegation require new mechanisms beyond the existing solutions within the principal-agent theory. Successful delegation in AI-to-human contexts requires a nuanced understanding of the interplay between human attitudes, AI characteristics, and the broader organizational and ethical landscape. Our results provide a first step into mitigating principal-agent problems arising from AI-to-human delegation. We contribute to research by reducing the uncertainty regarding the setting of autonomous AI artifacts possessing delegation ownership and giving instructions to humans.

5.6 Essay 6: Toward triadic delegation: How agentic IS artifacts affect the patient-doctor relationship in healthcare

In Essay 6, we explored the impact of agentic IS artifacts on the dyadic relationship between patient and doctor by conducting a single-case study on an AI-enabled agentic IS artifact designed to mitigate neurogenic bladder dysfunction.

Our findings reveal significant changes in agent attributes and agentic interactions, highlighting the emergence of conflicts within the triadic delegation relationship. While investigating the changes in agents' attributes, we identified novel attributes and interferences between the agents. Particularly, the attribute interferences provide agents with novel choices for delegating tasks within the triad. Moreover, we present novel interaction patterns facilitated by the agentic IS artifact. For instance, we could observe how the agentic IS artifact intervenes in the delegation between the two agents without being a proxy or delegator. Furthermore, our study points out the emergence of several conflicts arising from the triadic relationship, particularly around autonomy, information asymmetry, and role inference.

Moreover, we recognized that the novel agentic behavior and interaction enhances our theoretical understanding of triadic delegation, requiring theoretical embedding. Thus, we have captured the prevalent phenomena in our case and channeled them into a theoretical concept expanding existing delegation theory. Given the pivotal role of the agentic IS artifact in the patient-IS-doctor relationship, we propose that the agent relationship is more likely shaped in a sequential pattern instead of forming an equilateral triad. The delegation patterns mediation and moderation are particularly relevant in this sequential triadic delegation as it ensures the agentic IS artifact's involvement in human-to-human delegation through delegation-facilitating interventions. Beyond these delegation patterns, our results indicate that the agentic IS artifact also becomes increasingly involved in delegations. Its increasing capabilities enable it to perform tasks requiring high decision-making latitude and, therefore, executive more tasks.

Overall, our research contributes to theory and practice by offering a nuanced understanding of how agentic IS artifacts transform the patient-doctor relationship into a triadic one. We provide valuable insights into managing and optimizing the emergence of agentic IS artifacts, emphasizing the need to carefully consider the roles, capabilities, and interactions.

6 Discussion and conclusion

In the following, I summarize the results of my dissertation. In doing so, I critically appraise the key results of my dissertation and present their theoretical contributions and implications for practice. Lastly, I will reflect on my dissertation's limitations and outline promising avenues for future research.

6.1 Summary

Given the potential of autonomous AI agents to enhance efficiency and productivity across a wide range of activities, this dissertation aims to *guide the design and management of effective and beneficial collaboration between humans and autonomous AI agents*. Therefore, I postulated three research goals, allowing for a cumulative investigation of arising phenomena.

The first research goal of the dissertation focuses on providing an understanding of the design and management of AI applications, which I address through the first three essays of my dissertation. In Essay 1, I shed light on the relevant factors of managing AI applications and derive concrete AI management practices guiding organizations. Essay 2 complements my research on AI management by developing practices for monitoring ML applications as the prevailing technology concept enabling AI applications. The results of Essay 3 guide AI managers concerning the successful integration of AI services to enhance organizational AI capabilities. The second research goal was to expand the understanding of how the design modalities of AI agents' advice provisioning affect human-AI collaboration. Therefore, Essay 4 quantitatively evaluates the effect of timing and explanatory information on the mutual decision performance of human-AI teams. The research that I subsume under my third research goal, sheds light on the effects of increasing AI autonomy in human-AI collaboration. Accordingly, Essay 5 provides an understanding of how human-AI collaboration evolves when autonomous AI agents acquire task and delegation ownership in dyadic human-AI teams. Essay 6 expands the dyadic delegation relationship to a triadic delegation relationship, shedding light on how human-human interaction evolves through the integration of autonomous AI agents.

In summary, my dissertation enhances our understanding of human-AI collaboration through exploratory, prescriptive, descriptive, and analytical research, allowing for a comprehensive evaluation of the implications of integrating autonomous AI into organizational contexts.

6.2 Contributions to theory and implications for practice

Each essay in my dissertation offers valuable theoretical contributions and practical implications for designing and managing human-AI collaboration considering increasingly autonomous AI.

In addressing the first research goal, Essay 1 contributes to theory by proposing a theoretical framework for managing AI applications from an organizational perspective. The management model captures management factors and practices that promote information processing capabilities among AI stakeholders. AI applications' increasing intelligence and autonomy further drive AI's opacity, which in turn lever AI stakeholders' task uncertainty and equivocality (Berente et al., 2021). Therefore, the AI management model from Essay 1 enhances the theoretical understanding of AI management and, thus, expands organizational information processing theory by representing an instantiation of theoretical OIPT constructs. Essay 2 complements the contributions to the organizational perspectives on AI management by expanding the theoretical understanding of monitoring ML-based agents in their production environment. From a practical perspective, Essays 1 and 2 guide establishing organizational management and governance structures for successfully operating AI applications. Essay 3 further adds to the descriptive knowledge of organizational AI capability development. The proposed taxonomy represents a theoretical scaffolding structuring relevant properties of AI service offerings. The taxonomy expands the theoretical understanding of AI service platforms and advances digital platform theory in the context of AI service offerings. The taxonomy can guide organizations in developing sourcing and shoring strategies for AI capabilities.

Besides contributing to theory and practice on an organizational level, research goal 2 contributes to the interaction level. Through Essay 4, we increase the theoretical understanding of how the design modes of AI advice provisioning affect the decision accuracy of the human-AI system. Correspondingly, we further enhance the theoretical understanding of cognitive processes of human decision-making and the prevalence of human biases. By transferring our theoretical advancements into practice, our results help develop design guidelines for efficient human-AI interfaces that mitigate human biases and maximize decision accuracy. Implementing such design guidelines can enable greater hybrid intelligence and drive the persuasion of AI-enabled decision processes in practice.

In achieving research goal 3, my dissertation contributes to theory and practice by enhancing understanding of the effects of increasing AI autonomy on human-AI collaboration. The results of Essays 5 and 6 primarily contribute to the interaction and role levels. Essay 5 enhances conceptual understanding of increasing AI autonomy and its effects on the dyadic delegation relationship between superior AI agents and subordinate human agents. It has pushed the scientific frontier beyond the primarily human-driven task delegation theory by focusing on the increasing autonomy of AI agents. Moreover, as Essay 5 draws on the principal-agent theory, it also advances the theoretical understanding of the principal-agent relationship in AI-to-human delegation by examining and extending the concepts of the principal-agent theory. The practical insights help AI designers better navigate the complexities of integrating AI with high levels of autonomy into their workflows. By identifying key factors and requirements for effective AI-led delegation, this work provides actionable guidance for optimizing information flows and leveraging the complementary strengths of humans and AI agents. Essay 6 further contributes to delegation theory by conceptualizing the triadic delegation relationship between two human agents and one autonomous AI agent. Further, Essay 6 enhances delegation theory by describing the changing roles, interactions, and emerging conflicts in triadic delegations due to the AI agent's increased agency. Therefore, we theorize a new agency role —the mediator—as a dedicated role complementing the established roles of the delegator and the proxy. From a practical perspective, Essay 6 offers guidance for implementing autonomous AI in complex workflows, ensuring seamless integration, and avoiding conflicts such as agentic interferences. By theorizing the role of AI as a mediator, Essay 6 further provides actionable insights for AI designers to mitigate role conflicts and enhance collaboration efficiency. Additionally, it highlights the need for regulatory frameworks that address the increasing AI agency in human-AI collaboration.

Overall, the contributions of this dissertation underscore the transformative potential of autonomous AI in business and society. By advancing theoretical understanding across organizational, interactional, and role levels, the dissertation provides a comprehensive understanding of the complexities induced by AI autonomy. The practical insights offered equip practitioners with the necessary tools to design and manage AI systems that enhance collaboration and decision-making processes. As AI continues to evolve, this dissertation serves as a critical resource for researchers and practitioners, guiding the development and robust integration of autonomous AI agents.

6.3 Limitations

While this dissertation follows rigorous research standards and offers valuable insights into the design and management of human-AI collaboration, it is essential to acknowledge its limitations. Each essay in this dissertation presents unique constraints that collectively shape the boundaries of my research findings. Acknowledging these constraints offers critical insights for interpreting the results and can guide future research directions in human-AI collaboration. A comprehensive discussion of each essay's limitations is provided in their respective limitations sections. Consequently, I will now briefly outline the dissertation's primary overarching limitations, which refer to methodological and contextual constraints and conceptual limitations arising from AI's constantly evolving nature.

The methodological approaches in the essays of my dissertation mainly follow a qualitative research design focusing on a specific research case or context. While the single-case approach offers rich, contextual insights (Eisenhardt, 1989; Eisenhardt & Graebner, 2007), it also limits the robustness and generalizability of the derived theories. In Essays 1, 4, and 6, I contextualize my research within the healthcare domain, which is a rather complex and highly specialized environment, potentially restricting generalization to other contexts. Nonetheless, I recognize the healthcare sector's complexity and stringent regulatory requirements as a suitable field for studying human-AI collaboration. The critical challenges and rigorous standards in healthcare can mirror those in other high-stakes industries, such as finance, aviation, and cybersecurity. While rooted in healthcare, these similarities suggest that the essays' findings could also offer valuable insights into the dynamics in other domains where precision, reliability, and ethical considerations are equally paramount. Accordingly, I recommend multi-case studies or validation studies in the respective domains beyond healthcare to further mitigate the remaining constraints.

The second major limitation in my dissertation refers to a conceptual limitation arising from the rapidly evolving nature of AI technology and its autonomy. The theories and practical insights offered in this dissertation are based on the current state of AI technology. Particularly, Essays 5 and 6 approach phenomena associated with AI's increasing autonomy. In that regard, the rapid development of AI impedes capturing the full range of autonomy-induced effects on human-AI collaboration. As AI technology evolves, new phenomena may emerge requiring further exploration. The analysis of AI agents ascending to an equal footing with human agents and delegating tasks to humans marks a valuable first step in understanding the phenomena associated with AI's increasing autonomy.

6.4 Future research and outlook

This dissertation offers valuable insights into the collaboration between humans and autonomous AI agents. Alongside the limitations discussed, there are promising opportunities for future research in this emerging field.

Considering the methodological limitations of the dissertation's essays, I recommend future research to expand the contextual focus of my research by conducting multi-case studies or validation studies in domains beyond healthcare. Expanding research to various sectors will help validate the findings and improve the generalizability of the derived theories (Dubé & Paré, 2003; R. K. Yin, 2018). Particularly, sectors with similar advanced complexity and regulatory constraints can provide additional context for understanding the dynamics of human-AI collaboration in organizations. To address the limitations associated with AI's dynamic nature (Berente et al., 2021), I call for further research with organizations being at the forefront of integrating autonomous AI systems into their business processes. These organizations can offer rich insights into the practical challenges and successes of AI integration, aiding the refinement of theoretical models and frameworks. Moreover, engaging with industry practitioners through workshops, pilot projects, and collaborative platforms ensures that the research remains grounded in practical realities and adapts to technological advancements.

Furthermore, given that the implications associated with AI's autonomy affect not only the information systems realm, I call for more interdisciplinary research. Although information systems research already unifies social and technical sciences, expanding this collaboration to include fields such as ethics, law, psychology, and economics is crucial. This approach will enable more diverse perspectives, enhancing the robustness and relevance of future human-AI systems.

Beyond the scope of my dissertation, the evolving nature of AI presents numerous opportunities for future research that can build upon my findings. Considering the current state of AI and the latest research trends in the computer science domain (e.g., Händler, 2023; Weng, 2023; Wu et al., 2023), I anticipate increasing experimentation with the autonomous AI agent concept. In that regard, I call for dedicated information systems research, exploring the dynamics and efficiency of human-IS systems involving autonomous AI agents. Future research should focus on understanding how autonomous AI agents enable novel forms of business process automation and task augmentation while also addressing critical questions of responsibility and accountability. Additionally, research should focus on the changing nature of human tasks and required competencies, identifying how autonomous AI will alter these aspects and developing strategies to help the workforce adapt to new roles and skill requirements. Addressing these areas will provide a deeper understanding of the transformative impact of autonomous AI agents on business processes and human roles, guiding organizations in effectively integrating AI technologies to achieve productivity gains while ensuring ethical and responsible AI use.

In conclusion, this dissertation advances the understanding of human-AI collaboration in the context of increasing AI autonomy through exploratory, prescriptive, descriptive, and analytical research. The findings highlight the transformative potential of autonomous AI in business and society, offering practical insights for designing and managing AI systems that enhance collaboration and decision-making processes. As AI technology continues to evolve, this work serves as a critical resource for researchers and practitioners, guiding the development and robust integration of autonomous AI agents in organizations.

7 References

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE *Access : Practical Innovations, Open Solutions*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Adam, M., Diebel, C., Goutier, M., & Benlian, A. (2024). Navigating autonomy and control in human-AI delegation: User responses to technology- versus user-invoked task allocation. *DECISION SUPPORT SYSTEMS*, *180*, 114193. https://doi.org/10.1016/j.dss.2024.114193

Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction. *Journal of Economic Perspectives*, *33*(2), 31–50. https://doi.org/10.1257/jep.33.2.31

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software Engineering for Machine Learning: A Case Study. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)* (pp. 291–300). IEEE. https://doi.org/10.1109/ICSE-SEIP.2019.00042

Amershi, S., Inkpen, K., Teevan, J., Kikin-Gil, R., Horvitz, E., Weld, D [Dan], Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., & Bennett, P. N. (2019). Guidelines for Human-AI Interaction. In *Proceedings of CHI '19: CHI Conference on Human Factors in Computing Systems* (pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300233

Ananny, M., & Crawford, K. (2018). Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Arpteg, A., Brinne, B., Crnkovic-Friis, L., & Bosch, J. (2018). Software Engineering Challenges of Deep Learning. In *2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)* (pp. 50–59). IEEE. https://doi.org/10.1109/SEAA.2018.00018

Baier, L., Jöhren, F., & Seebacher, S. (2019). Challenges in the deployment and operation of machine learning in practice. *Proceedings of the 27th European Conference on Information Systems (ECIS)*, Article 163. https://aisel.aisnet.org/ecis2019_rp/163

Baird, A., & Maruping, L. M. (2021). The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts. *MIS Quarterly, 45*(1), 315–341. https://doi.org/10.25300/misq/2021/15882

Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T., & Weld, D [Daniel] (2021). Does the Whole Exceed its Parts? The Effect of AI Explanations on Complementary Team Performance. In *Proceedings of CHI '21: CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3411764.3445717

Benbya, H., Davenport, T. H., & Pachidi, S. (2020). Artificial Intelligence in Organizations: Current State and Future Opportunities. *MIS Quarterly Executive*, *19*(4), ix–xxi. https://doi.org/10.2139/ssrn.3741983

Benda, N. C., Novak, L. L., Reale, C., & Ancker, J. S. (2021). Trust in AI: why we should be designing for APPROPRIATE reliance. *Journal of the American Medical Informatics Association*, *29*(1), 207–212. https://doi.org/10.1093/jamia/ocab238

Benner, D., Schöbel, S., Janson, A., & Leimeister, J. M. (2022). How to achieve ethical persuasive design: A review and theoretical propositions for information systems. *AIS Transactions on Human-Computer Interaction*, *14*(4), 548–577. https://doi.org/10.17705/1thci.00179

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing Artificial Intelligence. *MIS Quarterly*, *45*(3), 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Bhattacherjee, A. (2012). *Social science research: Principles, methods, and practices*. *Open textbook library*. Anol Bhattacherjee and Open Textbook Library and Scholar Commons, University of South Florida.

Bondi, E., Koster, R., Sheahan, H., Chadwick, M., Bachrach, Y., Cemgil, T., Paquet, U., & Dvijotham, K. (2022). Role of Human-AI Interaction in Selective Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, *36*(5), 5286–5294. https://doi.org/10.1609/aaai.v36i5.20465

Bono, R., Alarcón, R., & Blanca, M. J. (2021). Report Quality of Generalized Linear Mixed Models in Psychology: A Systematic Review. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.666182

Borch, C. (2022). Machine learning, knowledge risk, and principal-agent problems in automated trading. *Technology in Society*, *68*(3), 101852. https://doi.org/10.1016/j.techsoc.2021.101852

Breck, E., Cai, S., Nielsen, E., Salib, M., & Sculley, D. (2017). The ML Test Score: A Rubric for ML Production Readiness and Technical Debt Reduction. *Proceedings of IEEE Big Data*.

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, *358*(6370), 1530–1534. https://doi.org/10.1126/science.aap8062

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What Can Machines Learn and What Does It Mean for Occupations and the Economy? *AEA Papers and Proceedings*, *108*, 43–47. https://doi.org/10.1257/pandp.20181019

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To Trust or to Think. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–21. https://doi.org/10.1145/3449287

Candrian, C., & Scherer, A. (2022). Rise of the machines: Delegating decisions to autonomous AI. *Computers in Human Behavior*, *134*, 107308. https://doi.org/10.1016/j.chb.2022.107308

Chao, C.-Y., Chang, T.-C., Wu, H.-C., Lin, Y.-S., & Chen, P.-C. (2016). The interrelationship between intelligent agents' characteristics and users' intention in a search engine by making beliefs and perceived risks mediators. *Computers in Human Behavior*, *64*, 117–125. https://doi.org/10.1016/j.chb.2016.06.031

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *Internation Journal of Information Management*, *60*, 102383. https://doi.org/10.1016/j.ijinfomgt.2021.102383

Corbin, J., & Strauss, A. (2015). *Basics of Qualitative Research: Techniques and procedures for developing grounded theory*. Sage Publications.

Cummings, M. (2017). Automation bias in intelligent time critical decision support systems. In D. Harris & W.-C. Li (Eds.), *Decision making in aviation* (pp. 289–294). Routledge.

Dattathrani, S., & De', R. (2023). The Concept of Agency in the Era of Artificial Intelligence: Dimensions and Degrees. *Information Systems Frontiers*, *25*(1), 29–54. https://doi.org/10.1007/s10796-022-10336-8

Davenport, T. H., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, *6*(2), 94–98. https://doi.org/10.7861/futurehosp.6-2-94

Davenport, T. H., & Ronanki, R. (2018). Artificial Intelligence for the Real World. *Harvard Business Review*, *96*(1), 108–116.

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid Intelligence. *Business & Information Systems Engineering*, *61*(5), 637–643. https://doi.org/10.1007/s12599-019-00595-2

Demetis, D., & Lee, A. (2018). When Humans Using the IT Artifact Becomes IT Using the Human Artifact. *Journal of the Association for Information Systems*, *19*(10). https://aisel.aisnet.org/jais/vol19/iss10/5

Dennis, A. R., Lakhiwal, A., & Sachdeva, A. (2023). AI Agents as Team Members: Effects on Satisfaction, Conflict, Trustworthiness, and Willingness to Work With. *Journal of Management Information Systems*, *40*(2), 307–337. https://doi.org/10.1080/07421222.2023.2196773

Dubé, & Paré (2003). Rigor in Information Systems Positivist Case Research: Current Practices, Trends, and Recommendations. *MIS Quarterly*, *27*(4), 597. https://doi.org/10.2307/30036550

Dung, L. (2024). Understanding Artificial Agency. *PHILOSOPHICAL QUARTERLY*. Advance online publication. https://doi.org/10.1093/pq/pqae010

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., . . . Williams, M. D. (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *Internation Journal of Information Management*, *57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *ACADEMY of MANAGEMENT REVIEW*, *14*(4), 532. https://doi.org/10.2307/258557

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory Building From Cases: Opportunities And Challenges. *Academy of Management Journal*, *50*(1), 25–32. https://doi.org/10.5465/amj.2007.24160888

Emirbayer, M., & Mische, A. (1998). What Is Agency? *American Journal of Sociology*, *103*(4), 962–1023. https://doi.org/10.1086/231294

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, *28*(1), 62–70. https://doi.org/10.1016/j.infoandorg.2018.02.005

Feulner, D., Guggenberger, T., Lämmermann, L., Protschky, D. (2024). Integrating artificial intelligence into football refereeing: Insights from German Bundesliga referees. *Working Paper*.

Fischer, V., & Beimborn, D. (2022). How Should Organizations Manage Artificial Intelligence? A Strategic Literature Review. *PACIS 2022 Proceedings*. https://aisel.aisnet.org/pacis2022/66

Fisher, G., Mayer, K., & Morris, S. (2021). From the Editors--Phenomenon-Based Theorizing. *ACADEMY of MANAGEMENT REVIEW*, *46*(4), 631–639. https://doi.org/10.5465/amr.2021.0320

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation. *Information Systems Research*, *33*(2), 678–696. https://doi.org/10.1287/isre.2021.1079

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), 678–696. https://doi.org/10.1287/isre.2021.1079

Galbraith, J. R. (1974). Organization design: An information processing view. *INTERFACES*, *4*(3), 28–36.

Garousi, V., Felderer, M., & Mäntylä, M. V. (2019). Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and Software Technology*, *106*, 101–121. https://doi.org/10.1016/j.infsof.2018.09.006

Gawer, A. (2009). Platform Dynamics and Strategies: From Products to Services. In A. Gawer (Ed.), *Platforms, Markets and Innovation* (pp. 45–76). Edward Elgar.

Gerlach, J. H., & Kuo, F.-Y. (1991). Understanding Human-Computer Interaction for Information Systems Design. *MIS Quarterly*, *15*(4), 527–549. https://doi.org/10.2307/249456

Gimpel, H., Gutheil, N., Mayer, V., Bandtel, M., Büttgen, M., Decker, S., Eymann, T., Feulner, S., Kaya, M. F., Kufner, M., Kühl, N., Lämmermann, L., Maedche, A., Ruiner, C., Schoop, M., & Urbach, N. (2024). *(Generative) AI Competencies for Future-Proof Graduates*. https://doi.org/10.5281/ZENODO.10680210

Gimpel, H., Hall, K., Decker, S., Eymann, T., Lämmermann, L., Braig, N., Maedche, A., Röglinger, M., Ruiner, C., Schoch, M., Schoop, M., Urbach, N., Vandirk, S., &

Gutheil, N. (2024). Using Generative AI in Higher Education – A Guide for Students and Lecturers. *Working Paper*.

Gimpel, H., Hall, K., Decker, S., Eymann, T., Lämmermann, L., Mädche, A., Röglinger, M., Ruiner, C., Schoch, M., Schoop, M., Urbach, N., & Vandirk, S. (2023). *Unlocking the Power of Generative AI Models and Systems such as GPT-4 and ChatGPT for Higher Education : A Guide for Students and Lecturers*. *Hohenheim Discussion Papers in Business, Economics and Social Sciences: Vol. 2023,02*. https://eref.uni-bayreuth.de/id/eprint/75892/

Gimpel, H., Hofbauer, S., Lämmermann, L., & Markgraf, M. (2024). How Should I Interact? - Interaction Style of AI-based Conversational Agents. *Working Paper*.

Gimpel, H., Hosseini, S., Huber, R., Probst, L., Röglinger, M., & Faisst, U. (2018). Structuring digital transformation: A framework of action fields and its application at ZEISS. *Journal of Information Technology Theory and Application (JITTA)*, *19*(1). https://aisel.aisnet.org/jitta/vol19/iss1/3

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking Qualitative Rigor in Inductive Research. *Organizational Research Methods*, *16*(1), 15–31. https://doi.org/10.1177/1094428112452151

Goldkuhl, G. (2012). Pragmatism vs interpretivism in qualitative information systems research. *European Journal of Information Systems*, *21*(2), 135–146. https://doi.org/10.1057/ejis.2011.54

Göndöcs, D., & Dörfler, V. (2024). Ai in medical diagnosis: Ai prediction & human judgment. *Artificial Intelligence in Medicine*, *149*, 102769. https://doi.org/10.1016/j.artmed.2024.102769

Gottschalk, C., Tewes, S., Niestroj, B. N., Jäger, C., Drees, J., & Ernst, A. (2022). Innovation in Elite Refereeing Through AI Technological Support for DOGSO Decisions. *International Journal of Operations Management*, *2*(3), 7–15. https://doi.org/10.18775/ijom.2757-0509.2020.23.4001

Gregor (2006). The Nature of Theory in Information Systems. *Management Information Systems Quarterly*, *30*(3), 611. https://doi.org/10.2307/25148742

Gregor, S., & Benbasat, I. (1999). Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice. *MIS Quarterly*, *23*(4), 497–530. https://doi.org/10.2307/249487

Gregory, R. W., & Henfridsson, O. (2021). Bridging Art and Science: Phenomenon-Driven Theorizing. *Journal of the Association for Information Systems*, *22*(6), 1509–1523. https://doi.org/10.17705/1jais.00703

Grover, V., & Lyytinen, K. (2015). New State of Play in Information Systems Research: The Push to the Edges. *Management Information Systems Quarterly*, *39*(2), 271–296. https://doi.org/10.25300/MISQ/2015/39.2.01

Händler, T. (2023, October 5). *Balancing Autonomy and Alignment: A Multi-Dimensional Taxonomy for Autonomous LLM-powered Multi-Agent Architectures*. http://arxiv.org/pdf/2310.03659

Harms, P. D., & Han, G. (2019). Algorithmic Leadership: The Future is Now. *Journal of Leadership Studies*, *12*(4), 74–75. https://doi.org/10.1002/jls.21615

Haußmann, C., Dwivedi, Y. K., Venkitachalam, K., & Williams, M. D. (2012). A summary and review of Galbraith's organizational information processing theory. In Dwivedi, Y., Wade, M., Schneberger, S. (Ed.), *29. Information Systems Theory: Integrated Series in Information Systems* (pp. 71–93). Springer, New York, NY. https://doi.org/10.1007/978-1-4419-9707-4_5

Hemmer, P., Schemmer, M., Kühl, N., Vössing, M., & Satzger, G. (2024, March 21). *Complementarity in Human-AI Collaboration: Concept, Sources, and Evidence*. http://arxiv.org/pdf/2404.00029v1

Hemmer, P., Schemmer, M., Vössing, M., & Kühl, N. (2021). Human-AI Complementarity in Hybrid Intelligence Systems: A Structured Literature Review. In *Proceedings of the 25th Pacific Asia Conference on Information Systems.* Association for Information Systems.

Hemmer, P., Westphal, M., Schemmer, M., Vetter, S., Vössing, M., & Satzger, G. (2023). Human-AI Collaboration: The Effect of AI Delegation on Human Task Performance and Task Satisfaction. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (pp. 453–463). ACM. https://doi.org/10.1145/3581641.3584052

Hirschheim, R., & Klein, H. (2012). A Glorious and Not-So-Short History of the Information Systems Field. *Journal of the Association for Information Systems*, *13*(4), 188–235. https://doi.org/10.17705/1jais.00294

Holst, L., Lämmermann, L., Mayer, V., & Urbach, N. (2024). The Impact of the EU AI Act's Transparency Requirements on AI Innovation. *Working Paper*.

Huang, M.-H., & Rust, R. T. (2018). Artificial Intelligence in Service. *Journal of Service Research, 21*(2), 155–172. https://doi.org/10.1177/1094670517752459

Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, *49*(1), 30–50. https://doi.org/10.1007/s11747-020-00749-9

Hummer, W., Muthusamy, V., Rausch, T., Dube, P., El Maghraoui, K., Murthi, A., & Oum, P. (2019). ModelOps: Cloud-based lifecycle management for reliable and trusted AI. In *2019 IEEE International Conference on Cloud Engineering (IC2E)* (pp. 113–120). IEEE. https://doi.org/10.1109/IC2E.2019.00025

Inkpen, K., Chappidi, S., Mallari, K., Nushi, B., Ramesh, D., Michelucci, P., Mandava, V., Vepřek, L. H., & Quinn, G. (2023). Advancing Human-AI Complementarity: The Impact of User Expertise and Algorithmic Tuning on Joint Decision Making. *ACM Transactions on Computer-Human Interaction, 30*(5), 1–29. https://doi.org/10.1145/3534561

The International Football Association Board. (2023). *Laws of the Game 23/24*. https://downloads.theifab.com/downloads/laws-of-the-game-2023-24?l=en

Jacobs, M., Pradier, M. F., McCoy, T. H., Perlis, R. H., Doshi-Velez, F., & Gajos, K. Z. (2021). How machine-learning recommendations influence clinician treatment selections: the example of the antidepressant selection. *Translational Psychiatry*, *11*(1), 108. https://doi.org/10.1038/s41398-021-01224-x

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007

Jensen, M. C., & Meckling, W. H. (1976). Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics*, 305–360.

Jiang, J., Kahai, S., & Yang, M. (2022). Who needs explanation and when? Juggling explainable AI and user epistemic uncertainty. *International Journal of Human-Computer Studies*, *165*. https://doi.org/10.1016/j.ijhcs.2022.102839

Jöhnk, J., Albrecht, T., Arnold, L., Guggenberger, T., Lämmermann, L., Schweizer, A., & Urbach, N. (2021). The Rise of the Machines: Conceptualizing the Machine Economy. *PACIS 2021 Proceedings*. https://aisel.aisnet.org/pacis2021/54

Jöhnk, J., Röglinger, M., Thimmel, M., & Urbach, N. (2017). How to Implement Agile IT Setups: A Taxonomy of Desgin Options. In I. Ramos, V. Tuunainen, & H. Krcmar (Chairs), *25th European Conference on Information Systems (ECIS).* Symposium conducted at the meeting of European Conference on Information Systems, Guimarães, Portugal.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). WHY ARE WE AVERSE TOWARDS ALGORITHMS? A COMPREHENSIVE LITERATURE REVIEW ON ALGORITHM AVERSION. *ECIS 2020 Research Papers*. https://aisel.aisnet.org/ecis2020_rp/168

Kim, E.-S. (2020). Deep learning and principal-agent problems of algorithmic governance: The new materialism perspective. *Technology in Society*, *63*(4), 101378. https://doi.org/10.1016/j.techsoc.2020.101378

Klaise, J., van Looveren, A., Cox, C., Vacanti, G., & Coca, A. *Monitoring and explainability of models in production*. http://arxiv.org/pdf/2007.06299v1

Köchling, A., Riazy, S., Wehner, M. C., & Simbeck, K. (2021). Highly Accurate, But Still Discriminatory. *Business & Information Systems Engineering*, *63*(1), 39–54. https://doi.org/10.1007/s12599-020-00673-w

Kreuzberger, D., Kühl, N., & Hirschl, S. (2023). Machine Learning Operations (MLOps): Overview, Definition, and Architecture. *IEEE Access : Practical Innovations, Open Solutions*, *11*, 31866–31879. https://doi.org/10.1109/ACCESS.2023.3262138

Krogh, G. von (2018). Artificial Intelligence in Organizations: New Opportunities for Phenomenon-Based Theorizing. *Academy of Management Discoveries*, *4*(4), 404–409. https://doi.org/10.5465/amd.2018.0084

Kühl, N., Schemmer, M., Goutier, M., & Satzger, G. (2022). Artificial intelligence and machine learning. *Electronic Markets*, *32*(4), 2235–2244. https://doi.org/10.1007/s12525-022-00598-0

Lai, V., Chen, C., Liao, Q. V., Smith-Renner, A., & Tan, C. *Towards a Science of Human-AI Decision Making: A Survey of Empirical Studies*. http://arxiv.org/pdf/2112.11471.pdf

Lämmermann, L., Richter, P., Zwickel, A., & Markgraf, M. (2022). AI fairness at subgroup level - A structured literature review. *ECIS 2022 Research Papers*. https://aisel.aisnet.org/ecis2022_rp/147

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction* (Second edition). Morgan Kaufmann Publishers, an imprint of Elsevier. https://learning.oreilly.com/library/view/-/9780128093436/?ar

Lee, A. S. (2001). Editor's Comments. *MIS Quarterly*, *25*(1), iii–vii. http://www.jstor.org/stable/3250954

Leyer, M., & Schneider, S. (2019). ME, YOU OR AI? HOW DO WE FEEL ABOUT DELEGATION. *Research Papers*. https://aisel.aisnet.org/ecis2019_rp/36

Li, J., Li, M., Wang, X., & Bennett Thatcher, J. (2021). Strategic directions for AI: The role of CIOs and boards of directors. *MIS Quarterly*, *45*(3), 1603–1644. https://doi.org/10.25300/MISQ/2021/16523

Li, T., Vorvoreanu, M., Debellis, D., & Amershi, S. (2023). Assessing Human-AI Interaction Early through Factorial Surveys: A Study on the Guidelines for Human-AI Interaction. *ACM Transactions on Computer-Human Interaction*, *30*(5), 1–45. https://doi.org/10.1145/3511605

Lins, S., Pandl, K. D., Teigeler, H., Thiebes, S., Bayer, C., & Sunyaev, A. (2021). Artificial Intelligence as a Service. *Business & Information Systems Engineering*, *63*(4), 441–456. https://doi.org/10.1007/s12599-021-00708-w

Lösser, B., Oberländer, A. M., & Rau, D. (2019). Taxonomy Research in Information Systems : A Systematic Assessment. In J. vom Brocke, S. Gregor, & O. Müller (Chairs), *27th European Conference on Information Systems (ECIS).* Symposium conducted at the meeting of European Conference on Information Systems (ECIS), Stockholm, Sweden.

Lyytinen, K., & King, J. (2004). Nothing At The Center? Academic Legitimacy in the Information Systems Field. *Journal of the Association for Information Systems*, *5*(6), 220–246. https://doi.org/10.17705/1jais.00051

Lyytinen, K., Nickerson, V, Jeffrey, & King, J. L. (2021). Metahuman systems = humans plus machines that learn. *JOURNAL of INFORMATION TECHNOLOGY*, *36*(4), 427–445. https://doi.org/10.1177/0268396220915917

Martinho, A., Kroesen, M., & Chorus, C. (2021). A healthy debate: Exploring the views of medical doctors on the ethics of artificial intelligence. *Artificial Intelligence in Medicine*, *121*, 102190. https://doi.org/10.1016/j.artmed.2021.102190

Mason, R. O., & Mitroff, I. I. (1973). A program for research on management information systems. *Management Science*, *19*(5), 475–487. https://doi.org/10.1287/mnsc.19.5.475

Mast, J. de, & Lokkerbol, J. (2012). An analysis of the Six Sigma DMAIC method from the perspective of problem solving. *International Journal of Production Economics*, *139*(2), 604–614. https://doi.org/10.1016/j.ijpe.2012.05.035

McCarthy, J., Minsky, M. L., Shannon, C. E., & Rochester, N. (1955). A proposal for the Dartmouth summer research project on artificial intelligence. Advance online publication. https://doi.org/10.1609/aimag.v27i4.1904

McCorduck, P., & Cfe, C. (2004). *Machines Who Think*. A K Peters/CRC Press. https://doi.org/10.1201/9780429258985

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2022). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, *54*(6), 1–35. https://doi.org/10.1145/3457607

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (Fourth edition). Jossey-Bass.

Mingers, J. (2001). Combining IS Research Methods: Towards a Pluralist Methodology. *Information Systems Research*, *12*(3), 240–259. http://www.jstor.org/stable/23011015

Murray, A., Rhymer, J., & Sirmon, D. G. (2021). HUMANS AND TECHNOLOGY: FORMS OF CONJOINED AGENCY IN ORGANIZATIONS. *ACADEMY of MANAGEMENT REVIEW*, *46*(3), 552–571. https://doi.org/10.5465/amr.2019.0186

Myers, M. D., & Newman, M. (2007). The Qualitative Interview in IS Research: Examining the Craft. *Information and Organization*, *17*(1), 2–26. https://doi.org/10.1016/j.infoandorg.2006.11.001

Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A Method for Taxonomy Development and its Application in Information Systems. *European Journal of Information Systems*, *22*(3), 336–359. https://doi.org/10.1057/ejis.2012.26

Ning, X., Lu, Y., Li, W., & Gupta, S. (2024). How transparency affects algorithmic advice utilization: The mediating roles of trusting beliefs. *DECISION SUPPORT SYSTEMS*, *183*. https://doi.org/10.1016/j.dss.2024.114273

Nserat, J., Braun, M., Kegel, F., & Kolbe, L. (2023). *57th Hawaii International Conference on System Sciences (HICCS), Waikiki Beach Resort, January 3-6, 2024*. *Proceedings of the Annual Hawaii International Conference on System Sciences (1999. Online)*. University of Hawaii at Manoa. https://scholarspace.manoa.hawaii.edu/items/14a09480-d58b-4c7d-8d30-85e4a25bfab8

Ogawa, R. T., & Malen, B. (1991). Towards Rigor in Reviews of Multivocal Literatures: Applying the Exploratory Case Study Method. *Review of Educational Research*, *61*(3), 265–286. https://doi.org/10.3102/00346543061003265

Orlikowski, W. J., & Iacono, C. S. (2001). Research Commentary: Desperately Seeking the "IT" in IT Research-A Call to Theorizing the IT Artifact. *Information Systems Research*, *12*(2), 121–134. https://doi.org/10.1287/isre.12.2.121.9700

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Peeters, M. M. M., van Diggelen, J., van den Bosch, K., Bronkhorst, A., Neerincx, M. A., Schraagen, J. M., & Raaijmakers, S. (2021). Hybrid collective intelligence in a human-AI society. *AI & Society, 36*(1), 217–238. https://doi.org/10.1007/s00146-020-01005-y

Phillips-Wren, G. (2013). Intelligent Decision Support Systems. In M. Doumpos & E. Grigoroudis (Eds.), *Multicriteria Decision Aid and Artificial Intelligence* (pp. 25–44). Wiley. https://doi.org/10.1002/9781118522516.ch2

Power, D. J. (2002). *Decision support systems: concepts and resources for managers*. Greenwood Publishing Group.

Puntoni, S., Reczek, R. W., Giesler, M., & Botti, S. (2021). Consumers and Artificial Intelligence: An Experiential Perspective. *Journal of Marketing*, *85*(1), 131–151. https://doi.org/10.1177/0022242920953847

Püschel, L., Schlott, H., & Röglinger, M. (2016). What's in a Smart Thing? Development of a Multi-layer Taxonomy. In P. J. Ågerfalk, N. Levina, & S. S. Kien (Chairs), *37th International Conference on Information Systems (ICIS).* Symposium conducted at the meeting of International Conference on Information Systems (ICIS), Dublin, Ireland.

Reverberi, C., Rigon, T., Solari, A., Hassan, C., Cherubini, P., & Cherubini, A. (2022). Experimental evidence of effective human-AI collaboration in medical decision-making. *Scientific Reports*, *12*(1), 14952. https://doi.org/10.1038/s41598-022-18751-2

Rudowsky, I. (2004). Intelligent Agents. *Communications of the Association for Information Systems*, *14*. https://doi.org/10.17705/1CAIS.01414

Russell, S., & Norvig, P. (2022). *Artificial intelligence: A modern approach*. Pearson Education Limited.

Saldaña, J. (2021). *The coding manual for qualitative researchers*. https://www.torrossa.com/gs/resourceproxy?an=5018667&publisher=fz7200

Sauerbrei, A., Kerasidou, A., Lucivero, F., & Hallowell, N. (2023). The impact of artificial intelligence on the person-centred, doctor-patient relationship: some problems and solutions. *BMC Medical Informatics and Decision Making*, *23*(1), 73. https://doi.org/10.1186/s12911-023-02162-y

Schemmer, M., Kuehl, N., Benz, C., Bartos, A., & Satzger, G. (2023). Appropriate Reliance on AI Advice: Conceptualization and the Effect of Explanations. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (pp. 410–422). Association for Computing Machinery. https://doi.org/10.1145/3581641.3584066

Schemmer, M., Kühl, N., Benz, C., & Satzger, G. (2022, April 19). *On the Influence of Explainable AI on Automation Bias*. http://arxiv.org/pdf/2204.08859

Schmitt, A., Zierau, N., Janson, A., & Leimeister, J. M. (2023). The Role of AI-Based Artifacts' Voice Capabilities for Agency Attribution. *Journal of the Association for Information Systems*, *24*(4), 980–1004. https://doi.org/10.17705/1jais.00827

Schoeffer, J., De-Arteaga, M., & Kuehl, N. (2022). *Explanations, Fairness, and Appropriate Reliance in Human-AI Decision-Making*. http://arxiv.org/pdf/2209.11812v3

Schuetz, S., & Venkatesh, V. (2020). *The Rise of Human Machines: How Cognitive Computing Systems Challenge Assumptions of User-System Interaction.*

Schultze, U., & Avital, M. (2011). Designing interviews to generate rich data for information systems research. *Information and Organization*, *21*(1), 1–16. https://doi.org/10.1016/j.infoandorg.2010.11.001

Shneiderman, B. (2020). Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human-Computer Interaction*, *36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Staw, B. M. (1981). The Escalation of Commitment to a Course of Action. *The Academy of Management Review*, *6*(4), 577–587. https://doi.org/10.2307/257636

Steffel, M., Williams, E. F., & Perrmann-Graham, J. (2016). Passing the buck: Delegating choices to others to avoid responsibility and blame. *Organizational Behavior and Human Decision Processes*, *135*, 32–44. https://doi.org/10.1016/j.obhdp.2016.04.006

Stelmaszak, Marta, Möhlmann, M., & Sørensen, C. (2024). When Algorithms Delegate to Humans: Exploring Human-Algorithm Interaction at Uber. *Management Information Systems Quarterly*.

Sturm, T., Gerlacha, J., Pumplun, L., Mesbah, N., Peters, F., Tauchert, C., Nan, N., & Buxmann, P. (2021). Coordinating Human and Machine Learning for Effective Organization Learning. *MIS Quarterly, 45*(3), 1581–1602. https://doi.org/10.25300/MISQ/2021/16543

Sujan, M., Furniss, D., Grundy, K., Grundy, H., Nelson, D., Elliott, M., White, S., Habli, I., & Reynolds, N. (2019). Human factors challenges for the safe use of artificial intelligence in patient care. *BMJ Health & Care Informatics*, *26*(1). https://doi.org/10.1136/bmjhci-2019-100081

Talone, A. (2019). The Effect of Reliability Information and Risk on Appropriate Reliance in an Autonomous Robot Teammate. *Electronic Theses and Dissertations*. https://stars.library.ucf.edu/etd/6852

Teodorescu, M., Morse, L., Awwad, Y., & Kane, G. (2021). Failures of fairness in automation require a deeper understanding of human-ML augmentation. *MIS Quarterly*, *45*(3), 1483–1500. https://doi.org/10.25300/MISQ/2021/16535

Terveen, L. G. (1995). Overview of human-computer collaboration. *Knowledge-Based Systems*, *8*(2-3), 67–81. https://doi.org/10.1016/0950-7051(95)98369-H

Thatcher, J., Pu, W., & Pienta, D. (2018). IS Information Systems a (Social) Science? *Communications of the Association for Information Systems*, 189–196. https://doi.org/10.17705/1CAIS.04311

Tonini, A., Mesquita Spinola, M., & Barbin Laurindo, F. (2006). Six Sigma and Software Development Process: DMAIC Improvements. In *2006 Technology Management for the Global Future - PICMET 2006 Conference* (pp. 2815–2823). IEEE. https://doi.org/10.1109/PICMET.2006.296875

Urbach, N., Albrecht, T., Guggenberger, T., Jöhnk, J., Arnold, L., Gebert, J., Jelito, D., Lämmermann, L., & Schweizer, A. (2020). *The Advance of the Machines: Vision und Implikationen einer Machine Economy*. https://eref.uni-bayreuth.de/id/eprint/58029/

Urbach, N., Häckel, B., Hofmann, P., Fabri, L., Ifland, S., Karnebogen, P., Krause, S., Lämmermann, L., Protschky, D., Markgraf, M., & Willburger, L. (2021). *KI-basierte Services intelligent gestalten: Einführung des KI-Service-Canvas*. https://eref.uni-bayreuth.de/id/eprint/66397/

van den Heuvel, W.-J., & Tamburri, D. A. (2020). Model-Driven ML-Ops for Intelligent Enterprise Applications: Vision, Approaches and Challenges. In B. Šiškov (Ed.), *Lecture Notes in Business Information Processing: Vol. 391. Business modeling and software design: 10th International Symposium, BMSD 2020, Berlin, Germany, July 6-8, 2020 : proceedings* (Vol. 391, pp. 169–181). Springer. https://doi.org/10.1007/978-3-030-52306-0\textunderscore

Vincent, V. U. (2021). Integrating intuition and artificial intelligence in organizational decision-making. *Business Horizons*, *64*(4), 425–438. https://doi.org/10.1016/j.bushor.2021.02.008

Vössing, M., Kühl, N., Lind, M., & Satzger, G. (2022). Designing Transparency for Effective Human-AI Collaboration. *Information Systems Frontiers*, *24*(3), 877–895. https://doi.org/10.1007/s10796-022-10284-3

Walsham, G. (1995). Interpretive case studies in IS research: nature and method. *European Journal of Information Systems*, *4*(2), 74–81. https://doi.org/10.1057/ejis.1995.9

Weber, M., Engert, M., Schaffer, N., Weking, J., & Krcmar, H. (2023). Organizational Capabilities for AI Implementation—Coping with Inscrutability and Data Dependency in AI. *Information Systems Frontiers*, *25*(4), 1549–1569. https://doi.org/10.1007/s10796-022-10297-y

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2), xiii–xxiii. http://www.jstor.org/stable/4132319

Wei, R., & Pardo, C. (2022). Artificial intelligence and SMEs: How can B2B SMEs leverage AI platforms to integrate AI technologies? *Industrial Marketing Management*, *107*, 466–483. https://doi.org/10.1016/j.indmarman.2022.10.008

Weng, L. (2023). LLM-powered Autonomous Agents. *Lil'log*. https://lilianweng.github.io/posts/2023-06-23-agent/

Wesche, J. S., & Sonderegger, A. (2019). When computers take the lead: The automation of leadership. *Computers in Human Behavior*, *101*(12), 197–209. https://doi.org/10.1016/j.chb.2019.07.027

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: service robots in the frontline. *Journal of Service Management*, *29*(5), 907–931. https://doi.org/10.1108/JOSM-04-2018-0119

Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: theory and practice. *The Knowledge Engineering Review*, *10*(2), 115–152. https://doi.org/10.1017/s0269888900008122

Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., Jiang, L., Zhang, X., Zhang, S., Liu, J., Awadallah, A. H., White, R. W., Burger, D., & Wang, C. (2023, August 16). *AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation*. http://arxiv.org/pdf/2308.08155

Xu, W., Dainoff, M. J., Ge, L., & Gao, Z. (2023). Transitioning to Human Interaction with AI Systems: New Challenges and Opportunities for HCI Professionals to Enable Human-Centered AI. *International Journal of Human-Computer Interaction*, *39*(3), 494–518. https://doi.org/10.1080/10447318.2022.2041900

Yin, J., Ngiam, K. Y., & Teo, H.-H. (2020). Work Design in Healthcare Artificial Intelligence Applications: The Role of Advice Provision Timing. In *Proceedings of the 41st International Conference on Information Systems.* Association for Information Systems.

Yin, R. K. (2018). *Case study research and applications: Design and methods* (Sixth edition). SAGE Publications, Inc.

Zhang, Q., Lee, M. L., & Carter, S. (2022). You Complete Me: Human-AI Teams and Complementary Expertise. In S. Barbosa, C. Lampe, C. Appert, D. A. Shamma, S. Drucker, J. Williamson, & K. Yatani (Eds.), *CHI Conference on Human Factors in Computing Systems* (pp. 1–28). ACM. https://doi.org/10.1145/3491102.3517791

Zhang, Y., Liao, Q. V., & Bellamy, R. K. E. (2020). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 295–305). Association for Computing Machinery. https://doi.org/10.1145/3351095.3372852

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), 1192–1213. https://doi.org/10.1287/isre.2021.1022

Zhu, H., Vigren, O., & Söderberg, I.-L. (2024). Implementing artificial intelligence empowered financial advisory services: A literature review and critical research agenda. *Journal of Business Research*, *174*, 114494. https://doi.org/10.1016/j.jbusres.2023.114494

Złotowski, J., Yogeeswaran, K., & Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies*, *100*, 48–54. https://doi.org/10.1016/j.ijhcs.2016.12.008

8 Appendix

8.1 Declarations of Co-Authorship and Individual Contributions

Essay 1: Managing artificial intelligence applications in healthcare: Promoting information processing among stakeholders

This essay is co-authored by three authors. I developed the research project. I was responsible for the conceptualization, theoretical grounding, data collection, formal analysis, and methodology of the research project. I also took the lead in writing the original draft and was responsible for the review and editing process. Thus, my co-authorship is reflected in the entire research project.

Essay 2: What gets measured gets improved: Monitoring machine learning applications in their production environments

This essay is co-authored by four authors. I contributed to methodological conceptualization and provided supervision throughout investigation and validation. I further participated in writing, reviewing and editing of the manuscript. Thus, my co-authorship is reflected in the entire research project.

Essay 3: Gateways to artificial intelligence: Developing a taxonomy for AI service platforms

This essay is co-authored by five authors. I contributed to data analysis and data validation of the research project. Moreover, I contributed to writing, reviewing and editing of the manuscript. Thus, my co-authorship is reflected in the entire research project.

Essay 4: Improving decision accuracy in human-AI collaboration: The role of timing and explanatory information

This essay is co-authored by five authors. I managed the project. I contributed to the conceptualization, investigation, data collection, formal analysis, and validation of the research. I wrote the original draft and participated in the review and editing of the manuscript. Thus, my co-authorship is reflected in the entire research project.

Essay 5: Task delegation from AI to humans: A principal-agent perspective

This essay is co-authored by five authors. I co-developed the research project contributed to the research design, investigation, and validation. I managed the project, provided supervision, and participated in the review and editing of the manuscript. Further, I participated in the research discussions. Thus, my co-authorship is reflected in the entire research project.

Essay 6: Toward triadic delegation: How agentic IS artifacts affect the patient-doctor relationship in healthcare

This essay is co-authored by five authors. I co-initiated and co-managed the project. I contributed to the conceptualization, investigation, data collection, data curation, formal analysis and validation of the research. I co-developed the original draft and participated in the review and editing of the manuscript. Further, I participated in the research discussions. Thus, my co-authorship is reflected in the entire research project.

8.2 Overview of other research articles

Table 2. *Overview of other published articles*

Reference	Title	Publication Outlet	VHB-2024
Holst et al. (2024)	The Impact of the EU AI Act's Transparency Requirements on AI Innovation	19th International Conference on Wirtschaftsinformatik	B
Gimpel, Gutheil, et al. (2024)	(Generative) AI Competencies for Future-Proof Graduates	Whitepaper	n/a
Gimpel et al. (2023)	Unlocking the power of generative AI models and systems such as GPT-4 and ChatGPT for higher education: A guide for students and lecturers	Whitepaper	n/a
Lämmermann et al. (2022)	AI Fairness at Subgroup Level - A Structured Literature Review	Proceedings of the European Conference on Information Systems (2022)	A
Urbach et al. (2021)	KI-basierte Services intelligent gestalten: Einführung des KI-Service-Canvas	Whitepaper	n/a
Jöhnk et al. (2021)	The Rise of the Machines: Conceptualizing the Machine Economy	Proceedings of the Pacific Asia Conference on Information Systems (2021)	C
Urbach et al. (2020)	The Advance of the Machines: Vision und Implikationen einer Machine Economy	Whitepaper	n/a

Table 3. *Overview of other unpublished articles*

Authors	Title	Publication Outlet & Status	VHB-2024
Feulner, Guggenberger, et al. (2024)	Integrating Artificial Intelligence into Football Refereeing: Insights from German Bundesliga Referees	European Conference on Information Systems (2025) *Status: Under Review*	A
Gimpel, Hall, et al. (2024)	Using Generative AI in Higher Education – A Guide for Students and Lecturers	Journal of Information Systems Education (JISE) *Status: Accepted*	C
Gimpel, Hofbauer, et al. (2024)	How Should I Interact? - Interaction Style of AI-based Conversational Agents	Human-Computer Interaction *Status: Under review*	B

Essay 1: Managing artificial intelligence applications in healthcare: Promoting information processing among stakeholders[2]

Authors

Lämmermann, Luis; Hofmann, Peter; Urbach, Nils

Abstract

AI applications hold great potential for improving healthcare. However, successfully operating AI is a complex endeavor requiring organizations to establish adequate management approaches. Managing AI applications requires functioning information exchange between a diverse set of stakeholders. Lacking information processing among stakeholders increases task uncertainty, hampering the operation of AI applications. Existing research lacks an understanding of holistic AI management approaches. To shed light on AI management in healthcare, we conducted a multi-perspective literature analysis followed by an interview study. Based on the organizational information processing theory, this paper investigates AI management in healthcare from an organizational perspective. As a result, we develop the AI application management model (AIAMA) that illustrates the managerial factors of AI management in healthcare and its interrelations. Furthermore, we provide managerial practices that improve information processing among stakeholders. We contribute to the academic discourse by providing a conceptual framework that increases the theoretical understanding of AI's management factors and understanding of management interrelations. Moreover, we contribute to practice by providing management practices that promote information processing and decrease task uncertainty when managing AI applications in healthcare.

Keywords: Artificial intelligence, Healthcare, Managing AI, Management model, Information processing

[2] This essay has been published in: Lämmermann, L., Hofmann, P., & Urbach, N. (2024). Managing artificial intelligence applications in healthcare: Promoting information processing among stakeholders. In *International Journal of Information Management*, 75, 102728

Essay 2: What gets measured gets improved: Monitoring machine learning applications in their production environments[3]

Authors

Protschky, Dominik; Lämmermann, Luis; Hofmann, Peter; Urbach, Nils

Abstract

Machine learning (ML) applications face many new, hardly predictable aspects in their production environments. Detecting new aspects in an ML production environment and understanding their impacts on the ML application is crucial if organizations are to ensure ML applications functionality. A monitoring entity is essential if one is to monitor ML applications in their production environments, to both continually minimize risks and improve ML application's performance. But existing monitoring approaches are struggling to deal with specifics that arise from ML applications. We aim at deriving monitoring practices and providing a holistic view over required steps in successful ML applications monitoring. Since there has been little research on this topic, we followed a qualitative research approach, i.e., we conducted an interview study combined with a multivocal literature review. Thus, we provide a theoretical framework of an ML-enabled agent in its production environment, five characteristics of ML applications' production environments and 17 monitoring practices – 14 practices arranged sequentially on a typical quality management cycle and three cross-sectional practices. To outline the ML specifics that arise in monitoring ML applications, we investigate the five ML production environment characteristics' influences on the ML monitoring practices.

Keywords: Monitoring, Production, Organizations, Measurement, Intelligent agents, Systematic literature review, Predictive models, Interviews, Data visualization, Training data

[3] This essay has been published in: Protschky, D., Lämmermann, L., Hofmann, P., & Urbach, N. (2025). What Gets Measured Gets Improved: Monitoring Machine Learning Applications in their Production Environments. In *IEEE Access*, 13, 34518-34538

Essay 3: Gateways to artificial intelligence: Developing a taxonomy for AI service platforms[4]

Authors

Geske, Flora; Hofmann, Peter; Lämmermann, Luis; Schlatt, Vincent; Urbach, Nils

Abstract

Artificial Intelligence (AI) carries the potential to drive innovation in many parts of today's business environment. Instead of building AI capabilities in-house, some organizations turn towards an emergent phenomenon: AI service platforms. However, as a novel concept in both research and practice, a systematic characterization of AI service platforms is missing. To address this gap, we define the concept of AI service platforms and develop a comprehensive taxonomy. Therefore, we rely on existing literature, 14 expert interviews, and a sample of 31 AI service platforms. Our contribution is threefold: First, our taxonomy systematically structures essential properties of AI service platforms, guiding future research and management practice. Second, we derive three generic motives of AI service platforms. Third, we contribute to the literature by critically discussing to what extent AI service platforms fit into the existing academic discourse on digital platforms and elaborate on future research directions.

Keywords: AI service platform, Digital platform, Artificial intelligence, Taxonomy

[4] This essay has been published in: Geske, F., Hofmann, P., Lämmermann, L., Schlatt, V., & Urbach, N. (2021). Gateways to Artificial Intelligence: Developing a Taxonomy for AI Service Platforms. In *ECIS 2021 Research Papers*, 150

Essay 4: Improving decision accuracy in human-AI collaboration: The role of timing and explanatory information

Authors

Gimpel, Henner; Lämmermann, Luis; Markgraf, Moritz; Urbach, Nils

Extented Abstract

Decision accuracy is vital in critical professional domains, such as healthcare, finance, refereeing, and legal adjudication, where flawed decisions can have significant consequences (Christensen & Knudsen, 2010; March, 1991; Simon, 1979). Organizations face increasing challenges due to the complexity and uncertainty inherent in modern decision-making environments, further exacerbated by human cognitive limitations, stress, and information overload (Adya & Phillips-Wren, 2020; Goddard et al., 2012). To mitigate these issues, Artificial Intelligence (AI) systems are emerging as powerful tools to support humans with decision-making in complex, uncertain, and ambiguous situations, owing to their analytical capabilities, speed and lack of cognitive biases (Benbya et al., 2020; Davenport & Kalakota, 2019; Huang & Rust, 2018; Jarrahi, 2018; Rastogi, 2023).

However, despite the analytical advantages of AI, the risks associated with AI making mistakes and the lack of human qualities such as intuition and contextual understanding prevent AI from fully automating decision making, especially in high-stakes situations. (Förster et al., 2020; Hemmer et al., 2021; V. Lai et al., 2021; Reis et al., 2020). As a result, many organizations adopt hybrid intelligence systems, combining human expertise and AI support within Decision Support Systems (DSS), where humans retain final decision authority (Dellermann, Ebel, et al., 2019; Hemmer et al., 2021). Effective human-AI collaboration requires careful consideration of socio-technical factors in interaction design to avoid biases such as automation bias and escalation of commitment, which can lead to inappropriate reliance on AI recommendations (Buçinca et al., 2021; Cummings, 2017; Schemmer et al., 2023).

Existing research indicates that timing and explanatory information significantly influence the effectiveness of interaction between AI-based DSS and humans. Providing AI advice prior to human's initial assessment may foster automation bias, leading humans to overly rely on AI recommendations without thorough evaluation (Cummings, 2017; Parasuraman & Riley, 1997). Conversely, providing AI advice to after initial human judgment might cause escalation of commitment, where humans disregard AI input by justifing their previous decisions rather than critically evaluating the situation (Staw, 1981). Additionally, explanatory information

provided with the AI advice can impact human reliance on AI advice (Bansal et al., 2021; Buçinca et al., 2021; Jacobs et al., 2021; Zhang et al., 2020). So far, there have been only a few quantitative investigations of the effect of timing and explanatory information on decision-accuracy. To address the uncertainty concerning their effects on AI advice provisioning in human-AI collaboration, we ask:

How do the timing and the explanatory information of an AI-based DSS advice affect decision accuracy?

To answer our research question, we conducted an experiment in which experienced football referees assessed real-world game situation. We examined decision accuracy under different conditions with respect to the timing of AI advice (before versus after human initial assessment) and their complementation with additional explanatory information. Our results show that AI-based DSS providing decision support after human assessment combined with explanatory information can significantly improve decision accuracy of the human-AI team.

Our findings advance the theoretical understanding of human-AI interaction design and provide practical insights for improving the decision accuracy of human-AI collaboration, highlighting the relevance of considering timing and explanatory information of advice from AI-based DSS.

References

Adya, M., & Phillips-Wren, G. (2020). Stressed decision makers and use of decision aids: a literature review and conceptual model. *Information Technology & People*, *33*(2), 710–754. https://doi.org/10.1108/ITP-04-2019-0194

Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T., & Weld, D [Daniel] (2021). Does the Whole Exceed its Parts? The Effect of AI Explanations on Complementary Team Performance. In *Proceedings of CHI '21: CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3411764.3445717

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To Trust or to Think. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), Article 188, 1–21. https://doi.org/10.1145/3449287

Christensen, M., & Knudsen, T. (2010). Design of Decision-Making Organizations. *Management Science*, *56*(1), 71–89. https://doi.org/10.1287/mnsc.1090.1096

Cummings, M. (2017). Automation bias in intelligent time critical decision support systems. In D. Harris & W.-C. Li (Eds.), *Decision making in aviation* (pp. 289–294). Routledge.

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid Intelligence. *Business & Information Systems Engineering*, *61*(5), 637–643. https://doi.org/10.1007/s12599-019-00595-2

Förster, M., Klier, M., Kluge, K., & and Sigler, I. (2020). Evaluating explainable Artifical intelligence – What users really appreciate. In *Proceedings of the 28th European Conference on Information Systems.* Association for Information Systems.

Hemmer, P., Schemmer, M., Vössing, M., & Kühl, N. (2021). Human-AI Complementarity in Hybrid Intelligence Systems: A Structured Literature Review. In *Proceedings of the 25th Pacific Asia Conference on Information Systems.* Association for Information Systems.

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007

Lai, V., Chen, C., Liao, Q. V., Smith-Renner, A., & Tan, C. (2021). *Towards a Science of Human-AI Decision Making: A Survey of Empirical Studies*. http://arxiv.org/pdf/2112.11471.pdf

March, J. (1991). How Decisions Happen in Organizations. *Human-Computer Interaction*, *6*(2), 95–117. https://doi.org/10.1207/s15327051hci0602_1

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Rastogi, C. (2023). Investigating the Relative Strengths of Humans and Machine Learning in Decision-Making. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 987–989). Association for Computing Machinery. https://doi.org/10.1145/3600211.3604738

Reis, L., Maier, C., Mattke, J., Creutzenberg, M., & Weitzel, T. (2020). Addressing User Resistance Would Have Prevented a Healthcare AI Project Failure. *MIS Quarterly Executive*, *19*(4), Article 8, 279–296. https://doi.org/10.17705/2msqe.00038

Schemmer, M., Kuehl, N., Benz, C., Bartos, A., & Satzger, G. (2023). Appropriate Reliance on AI Advice: Conceptualization and the Effect of Explanations. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (pp. 410–422). Association for Computing Machinery. https://doi.org/10.1145/3581641.3584066

Simon, H. A. (1979). Rational Decision Making in Business Organizations. *The American Economic Review*, *69*(4), 493–513. https://www.jstor.org/stable/1808698

Staw, B. M. (1981). The Escalation of Commitment to a Course of Action. *The Academy of Management Review*, *6*(4), 577–587. https://doi.org/10.2307/257636

Essay 5: Task delegation from AI to humans: A principal-agent perspective[5]

Authors

Guggenberger, Tobias; Lämmermann, Luis; Urbach, Nils; Walter, Anna; Hofmann, Peter

Abstract

Increasingly intelligent AI artifacts in human-AI systems perform tasks more autonomously as entities that guide human actions, even changing the direction of task delegation between humans and AI. It has been shown that human-AI systems achieve better results when the AI artifact takes the leading role and delegates tasks to a human rather than the other way around. This study presents phenomena, conflicts, and challenges that arise in this process, explored through the theoretical lens of principal-agent theory (PAT). The findings are derived from a systematic literature review and an exploratory interview study and are placed in the context of existing constructs of PAT. Furthermore, this article paper identifies new causes of tensions that arise specifically in AI-to-human delegation and calls for special mechanisms beyond the classical solutions of PAT. The paper thus contributes to the understanding of autonomous AI and its implications for human-AI delegation.

Keywords: Delegation, artificial intelligence, human-AI collaboration, principal-agent theory

[5] This essay has been published in: Guggenberger, T., Lämmermann, L., Urbach, N., Walter, A. M., & Hofmann, P. (2023). Task delegation from AI to humans: a principal-agent perspective. In *ICIS 2023 Proceedings*, 13

Essay 6: Toward triadic delegation: How agentic IS artifacts affect the patient-doctor relationship in healthcare

Authors

Fechner, Pascal; Lämmermann, Luis; Lockl, Jannik; Röglinger, Maximilian; Urbach, Nils

Extented Abstract

The rise of agentic information systems (IS) in healthcare marks a significant shift in the patient-doctor relationship, as these systems increasingly exhibit autonomous behavior and decision-making latitude. Traditionally, IS artifacts were seen as passive tools assisting human agents (Baird & Maruping, 2021; Orlikowski & Iacono, 2001). However, with advances in artificial intelligence (AI), they are now capable of delegating tasks, transferring both rights and responsibilities, and actively shaping decision-making processes (Baird & Maruping, 2021; Berente et al., 2021). This shift leads to a transformation of the previously dyadic relationship between patient and doctor, introducing a triadic relationship with the patient, the doctor and an agentic IS artifact on equal footing.

Despite the growing persuasion of agentic IS artifacts in healthcare, existing research does not fully account for the transition from a dyadic patient-doctor relationship to a triadic relationship that includes an agentic IS artifact. Understanding this transition is critical to ensuring that agentic IS artifacts enhance, rather than disrupt, healthcare delivery. Therefore, we ask:

How do agentic IS artifacts affect the dyadic patient-doctor relationship in patient-centric healthcare delivery?

Drawing on phenomenon-based theorizing (Fisher et al., 2021; Gregory & Henfridsson, 2021), we conducted an exploratory single-case study (Eisenhardt & Graebner, 2007; Lee, 1989) focused on an agentic health companion supporting patients with neurogenic lower urinary tract dysfunction. We examine this transformation through the theoretical lens of delegation theory (Baird & Maruping, 2021), highlighting how the emergence of agentic IS artifacts alters delegation structures, agentic roles, and decision-making authority in the agentic triad.

Our findings reveal significant changes in agent attributes and interactions, also highlighting emerging conflicts in triadic delegation. We identify novel attributes and attributes interferences that provide agents with novel delegation choices. Additionally, we uncover novel interaction patterns, including the agentic IS artifact's role in intervening in patient-doctors interactions without being a direct proxy or delegator. Key conflicts arise around autonomy, information

asymmetry, and role interference, underscoring the complexity of integrating agentic IS artifacts into healthcare. Our study further suggests that rather than forming an equilateral triad, the patient-IS-doctor relationship follows a sequential pattern, as the artifact increasingly engages in delegation, executing tasks with greater decision-making latitude. Moreover, the agentic IS facilitates human-to-human delegation through mediation and moderation patterns.

Our study makes two main contributions: Theoretically, it expands our understanding of how agentic IS artifacts transform the traditionally dyadic patient-doctor relationship into a triadic delegation structure. By examining how agentic IS artifacts acquire rights and responsibilities, we illustrate that their increased decision-making latitude can place them on equal footing with human agents, necessitating a new theoretical framework of triadic delegation. Practically, our findings emphasize the need to design and regulate agentic IS artifacts with careful consideration of their roles, boundaries, and interactions. Designers must anticipate potential conflicts, such as role interference and autonomy loss, and implement mechanisms that preserve shared decision-making.

References

Baird, A., & Maruping, L. (2021). The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts. *Management Information Systems Quarterly*, *45*(1), 315–341. https://aisel.aisnet.org/misq/vol45/iss1/12

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing Artificial Intelligence. *MIS Quarterly*, *45*(3), 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory Building From Cases: Opportunities And Challenges. *Academy of Management Journal*, *50*(1), 25–32. https://doi.org/10.5465/amj.2007.24160888

Fisher, G., Mayer, K., & Morris, S. (2021). From the Editors—Phenomenon-Based Theorizing. *The Academy of Management Review*, *46*(4), 631–639. https://doi.org/10.5465/amr.2021.0320

Gregory, R. W., & Henfridsson, O. (2021). Bridging Art and Science: Phenomenon-Driven Theorizing. *Journal of the Association for Information Systems*, *22*(6), 1509–1523. https://doi.org/10.17705/1jais.00703

Lee, A. S. (1989). A Scientific Methodology for MIS Case Studies. *MIS Quarterly*, *13*(1), 33. https://doi.org/10.2307/248698

Orlikowski, W. J., & Iacono, C. S. (2001). Research Commentary: Desperately Seeking the "IT" in IT Research—A Call to Theorizing the IT Artifact. *Information Systems Research*, *12*(2), 121–134. https://doi.org/10.1287/isre.12.2.121.9700