Teaming with AI: Human Behavior in AI-Assisted Decision-Making

der Fakultät für Wirtschaftswissenschaften
der Universität Paderborn

zur Erlangung des akademischen Grades
Doktor der Wirtschaftswissenschaften
- Doctor rerum politicarum –

vorgelegte Dissertation
von
Jörg Papenkordt
geboren am 21. Dezember 1994 in Paderborn

Paderborn 2025

Dekan	Prof. Dr. Jens Müller
Gutachter	Prof. Dr. Kirsten Thommes Prof. Dr. Philip Yang

Acknowledgements

When I began the journey of my doctoral studies, I was not aware of what it required of me personally, mentally, and academically, and, more importantly, what it enabled me to achieve. This cumulative doctoral thesis would not have been possible without the dedication, endless support, and motivation of many amazing people.

First, I am more than grateful to my entire family and all my friends for always inspiring and supporting me. In particular, I would like to thank my fiancée, Eileen Obergassel, for her limitless belief in me and her unconditional support throughout this journey. Without your constant love, this dissertation would not have been possible. In addition, I would like to express my special gratitude to my parents, Johannes and Roswitha Papenkordt, for paving the way for this career path and continuing to encourage me throughout my entire academic endeavor. Additionally, I am very grateful to my older brother, Stefan Papenkordt, his wife, Jana, and their beautiful little daughter, Frieda, for providing me with the stability I needed. Finally, I want to thank my brother for always motivating me.

Furthermore, my special thanks go to my supervisor, Prof. Dr. Kirsten Thommes, for her invaluable guidance and support in navigating the complexities of academic research. Her commitment, expertise, and perseverance inspired me to strive for higher standards, which made this thesis possible. Being part of her chair encouraged both my professional and personal growth by consistently challenging me. I would also like to thank all my colleagues and co-authors, particularly Jana Kim Gutt, Miro Mehic, Jaroslaw Kornowicz, Olesja Lammert, Anastasia Lebedeva, Axel-Cyrille Ngonga Ngomo, Johannes Dahlke, Nicolas Neef, and Sarah Zabel, for the fruitful and constructive discussions during our research. Thank you!

Paderborn, July 2025 Jörg Papenkordt

Contents

List of Figures

Part A

Part B

List of Tables

Part A

Part B

List of Abbreviations

AI Artificial Intelligence

GenAI Generative Artificial Intelligence

HCI Human-Computer Interaction

I-P-O Input-Process-Output

IS Information Systems

LLM Large Language Model

SHAP SHapley Additive exPlanations

XAI Explainable Artificial Intelligence

Part A
Research Overview

1. Synopsis

> *"This new form of AI excels at modeling human intuition rather than human reasoning and it will enable us to create highly intelligent and knowledgeable assistants who will increase productivity in almost all industries. If the benefits of the increased productivity can be shared equally it will be a wonderful advance for all humanity. Unfortunately, the rapid progress in AI comes with many [...] risks."*
>
> *Geoffrey Hinton, Nobel Prize winner, 2024*

Artificial intelligence (AI) is revolutionizing and transforming our world, redefining how humans interact with machines (Ågerfalk, 2020; Grønsund and Aanestad, 2020; Raisch and Krakowski, 2021; Baird and Maruping, 2021; Millet et al., 2023; Dell'Acqua et al., 2023; Vaccaro et al., 2024; Raisch and Fomina, 2025; Brynjolfsson et al., 2025). As underscored by the speech of Hinton (2024) at the Nobel Prize banquet and reflected in numerous research studies, media reports, policy initiatives, business investments, and legislation, AI presents both novel opportunities and complex challenges for humanity. While recent research has predominantly focused on technological advancements, it often overlooks the human-user perspective and its impact on the interaction with AI systems (Wilder et al., 2020; Schuetz and Venkatesh, 2020; van der Waa et al., 2021; Raisch and Krakowski, 2021; Baird and Maruping, 2021; Murray et al., 2021).

Unlike previous technologies, AI-driven systems exhibit distinctive traits—such as their probabilistic and often opaque nature, their capacity to mimic human-like behavior, and potential to surpass human intelligence—which make human-AI interaction particularly demanding (Miller, 2019; Arrieta et al., 2020; Jussupow et al., 2020; Amabile, 2020; Vilone and Longo, 2021; Fügener et al., 2022; Vaccaro et al., 2024; Larson et al., 2024; Brynjolfsson et al., 2025). These characteristics raise fundamental questions that current human-computer interaction (HCI) frameworks often struggle to address (Von Krogh, 2018; Ågerfalk, 2020; Yang et al., 2020; Schuetz and Venkatesh, 2020; Baird and Maruping, 2021; Fügener et al., 2021; O'Neill et al., 2023). Accordingly, there is a broad consensus that existing HCI guidelines must be extended to fully harness the capabilities of this revolutionary technology.

This dissertation contributes to closing this research gap by investigating how individuals and teams interact with AI, with a focus on the socio-cognitive factors that shape decision-making and collaboration. At its core, this dissertation argues that while AI systems increasingly replicate and, in some cases, exceed human cognitive abilities, their success often depends on how human decision-makers perceive, understand, and engage with AI-generated recommendations. Despite progress in human-AI research, current approaches frequently neglect users' varying expertise levels, cognitive limitations, and the social context of collaboration. This work addresses these limitations through four key research questions, each examining a critical aspect of human-AI interaction: explanation modality, explanation accessibility, algorithm aversion, and team collaboration dynamics. By integrating theoretical insights from multiple disciplines, this thesis aims to advance our understanding of how decision-making and collaboration unfold in human-AI interaction.

1.1. Introduction

AI remains an ambiguous concept despite its growing ubiquity. AI has become a buzzword encompassing various technologies, leading to definitional inconsistencies (Burrell, 2016; Ågerfalk, 2020). Conventional definitions tend to highlight the technical characteristics of AI while overlooking its human-centric purpose (Caluori, 2024). Alan Turing's foundational but simple test introduced a widely accepted notion of AI, proposing that if a human interlocutor within an interaction cannot distinguish between communicating with a machine or a human, the system can be considered intelligent (Turing, 2004). Nonetheless, this definition lacks formal precision. As a result, AI remains an elusive concept, often obscured by technical ambiguity and differing interpretations. A more intuitive approach is to examine the definitions of "human intelligence" as a starting point. Wechsler (1958) defines human intelligence as the ability to act with purpose, reason logically, and interact efficiently with the environment. Compared to other technologies, the unique aspect of AI systems is their contextual, adaptive, and interactive nature (Schuetz and Venkatesh, 2020;

Ågerfalk, 2020; Baird and Maruping, 2021). This development signifies that AI systems are no longer passive or static; they can learn over time, respond dynamically to their environment, and interact with humans in a human-like manner (Schuetz and Venkatesh, 2020; Ågerfalk, 2020; Baird and Maruping, 2021).

Historically, the ability to process unstructured data, such as text, audio, and visual information, has been considered a uniquely human trait (Schuetz and Venkatesh, 2020; Guzman and Lewis, 2020; Argote et al., 2021; Dell'Acqua et al., 2023). AI has now been enabled to organize, categorize, and interpret these data, allowing it to enter domains previously exclusive to human cognition (Raisch and Fomina, 2025; Epstein et al., 2023; Noy and Zhang, 2023; Peng et al., 2023; Doshi and Hauser, 2024; Brynjolfsson et al., 2025; Dell'Acqua et al., 2025). This arises from AI models being frequently probabilistic instead of deterministic by moving their computation protocol from operating on explicit rules, such as if-then conditions, to training neural networks to select the most probable accurate answer (Schuetz and Venkatesh, 2020). In other words, the machine derives complex statistical models by identifying patterns within the data (Agrawal et al., 2017). Hence, AI introduces a significant shift in user-machine interactions, presenting numerous research questions that current theories have not yet addressed (Schuetz and Venkatesh, 2020; Ågerfalk, 2020; Yang et al., 2020; Baird and Maruping, 2021; O'Neill et al., 2023). One of the most notable expressions of this shift is that humans increasingly perceive AI systems as teammates rather than tools, shaping the user interaction with AI (Seeber et al., 2020; Wolf and Stock-Homburg, 2023; De Freitas et al., 2025; Dell'Acqua et al., 2025).

Humans frequently anthropomorphize by attributing human characteristics to non-human agents, such as objects and animals. This tendency is especially evident in human-AI interactions, where research shows that these interactions often mimic human-to-human exchanges rather than human-to-object ones, and individuals start to humanize AI systems (Seeber et al., 2020; Wan et al., 2024; De Freitas et al., 2025; Dell'Acqua et al., 2025). Therefore, it is logical to assume that humans will consciously or unconsciously adopt behaviors from human-to-human interaction and transfer them to human-AI interaction. Social cognitive theories postulate that individuals classify their surroundings, distinguishing between humans, animals, and objects (Kunda, 1999). In the realm of AI, this simplicity is lost, as evidenced by the Turing test, which demonstrates that humans often cannot reliably discern whether they are engaging with an AI or a fellow human (Warwick and Shah, 2016; Köbis and Mossink, 2021; Kovács, 2024). According to social response theory, humans react to technologies that exhibit human-like characteristics or social signals in a similar way as they would to other humans, even with the knowledge that they are engaging with a technology (Moon, 2000). Consequently, humans use fundamental social scripts when engaging with computers that display human-like traits or actions. This reasoning allows for testing and validating various interpersonal theories in human-AI interaction. The core of every interaction is the underlying trust (Thiebes et al., 2021). In human-to-human interaction, trust is based on a cost-benefit calculation (an economic perspective) or sociopsychological perceptions (Lewicki, 1996).

The cost-benefit principle emerged from behavioral decision-making research, emphasizing (cognitive) effort as a key factor in selecting decision-making strategies (Gregor and Benbasat, 1999). Cognitive effort refers to the basic information-processing activities required to complete a task (Johnson and Payne, 1985). Typically, a decision-maker evaluates an advisor's recommendation by weighing the potential benefits and costs (Bonaccio and Dalal, 2006). Humans generally avoid exerting effort to access and engage with explanations unless they perceive that the expected benefits of cognitive engagement outweigh the required mental effort (Gregor and Benbasat, 1999; Shenhav et al., 2017). Consequently, given a choice between similarly rewarding options, humans tend to prefer the associated option that minimizes their effort ("law of less work") (Hull, 1943). However, this cost-benefit evaluation is further influenced by the fact that decision-makers take other exogenous factors into account to precisely specify their rationale (such as their abilities or the abilities of the advisor in this area, the complexity of the task, experience with the advisor, etc.) (Bonaccio and Dalal, 2006). In human-AI interaction, it appears logical for humans to rely on AI recommendations that exceed expert capabilities, aiming to reduce cognitive effort and enhance benefits. From the perspective of Kahnemann's dual system theory, humans should ideally engage in deliberate and analytical thinking (System 2/ slow-thinking) rather than relying on intuition or gut feeling (System 1/ fast-thinking) when making decisions (Kahneman, 2011). However, applying social frameworks to the Information systems (IS) literature must acknowledge that humans do not always behave rationally and often rely on their feelings (Buçinca et al., 2021; Bonnefon and Rahwan, 2020; Kahneman et al., 2021). Humans frequently apply decision heuristics or folk theories when making decisions (Kahneman, 2011; Zedelius et al., 2017). Folk theories form expectations rooted in experience, yet not systematically verified (Rip, 2019). As a result, appraisals, such as expectations and feelings, are crucial in determining advice acceptance, algorithm utilization, and task delegation to AI systems (Baird and Maruping, 2021). Thus, collaboration hinges on the human perception of the AI system and a cognitive assessment of the delegation's cost-benefit ratio. Understanding the reasons, timing, and methods of human reliance on AI systems is crucial, as well as identifying elements that enhance or impair their interaction (Baird and Maruping, 2021).

In addition, the expanding capabilities of AI present pioneering opportunities and critical challenges. On the one hand, AI has demonstrated revolutionary performance in various domains, often exceeding human capabilities (Enholm et al., 2022). For example, AI-driven solutions are being applied in increasingly heterogeneous domains, including robotic surgery (Diana and Marescaux, 2015),[1] assisted driving (Hecker et al., 2018), hiring new personnel (Gil et al., 2020), personal and material logistics (Bader and Oevermann, 2017), finance (Strich et al., 2021), and even crime prediction (Završnik, 2020). On the other hand, AI's probabilistic nature also introduces challenges. Since AI models continuously learn and evolve, their behavior can become unpredictable and inconsistent (Ågerfalk, 2020). Even data that enters the system is often unclear or undisclosed (Burrell, 2016). Moreover, these models often make forecast mistakes, particularly for instances underrepresented in the training data set (Yang et al., 2020). The increasing complexity also frequently makes AI decision-making processes opaque to users, contributing to a lack of transparency, trust, and reliance (Burrell, 2016; Wang and Benbasat, 2016). Furthermore, AI applications have shown significant drawbacks, including discrimination (Tambe et al., 2019), the spread of misinformation (Rebholz et al., 2024), and safety risks in fields such as autonomous driving and robotic surgery (Varshney and Alemzadeh, 2017). These issues indicate that AI systems are certainly not without flaws. In particular, in the case of decisions that directly affect human individuals (e.g., credit scoring, insurance qualification, hiring), those affected have a right to know the reasons for the decision (Burrell, 2016; EU, 2021). Beyond legal and ethical considerations, technical and economic reasons also reveal that AI should collaborate with humans rather than act autonomously. In their seminal paper, Mittelstadt et al. (2019) distinguish between epistemic and normative concerns that pose challenges to the use of AI in practice. First, algorithms may result in inconclusive evidence due to statistical handling. Second, the relation between the data and the conclusions reached by an algorithm should be evident, but in reality, it is often inscrutable. Third, the data used may be too limited to assess new cases, resulting in misguided evidence. From

[1] 693,000 operations were performed in the US in 2017 alone.

a normative perspective, the results of an algorithm may be perceived as unfair. In addition, algorithms may, in turn, transform the way individuals perceive the world, which poses further ethical questions about their use. Finally, the root causes of harm cannot be easily traced when algorithms are involved. Based on these concerns, AI fails to carry out all activities independently, so collaboration with a human is required (Brynjolfsson et al., 2018; Brynjolfsson et al., 2025). This paradigm shift from automation to augmentation, in which humans learn from AI and vice versa, is further supported by the observation that humans and AI are more effective as a team than when acting alone (Schuetz and Venkatesh, 2020; Raisch and Krakowski, 2021).

Nevertheless, concerns about algorithmic bias, ethical dilemmas, or legal regulations boiled down to the "black-box" problem of AI systems, necessitating further research into explainable AI (XAI). In other words, opaqueness is the central concern in this context, referring to the lack of transparency in how AI processes input to generate output (Burrell, 2016). To address this issue, decision-makers from politics, industry, science, and society are highlighting the urgent need for AI and its decisions to become more transparent and understandable to users (Abdul et al., 2018; EU, 2021; Mohseni et al., 2021; Schemmer et al., 2022). Consequently, the XAI research domain seeks to elucidate the "black-box" nature of AI systems and explain the reasoning behind specific decisions or recommendations, thereby aiding users in comprehension, appropriate reliance, and effective management of the system (Gunning and Aha, 2019; Arrieta et al., 2020; Miller, 2019). In recent years, the topic of XAI has experienced a real boost in research. As XAI is a research field at the intersection of IS, HCI, and social sciences, the multidisciplinary consideration has led to the development of a wide range of techniques to explain the rationale behind AI decisions, as well as various taxonomies of these techniques (Wang and Yin, 2021). Currently, the research field suffers from some conceptual problems as terms like "explainability", "understandability", "transparency", "interpretability", "comprehensibility", or "trust" and "reliance" are frequently used synonymously even though they are not interchangeable. As a result, research comes to seemingly contradictory conclusions, even though different influencing factors have been considered (Schemmer et al., 2022). To prevent misunderstandings, the terms are differentiated in this thesis as follows:

- Explainability is related to the notion of explanation as an interface between humans and an AI system (Arrieta et al., 2020; Gilpin et al., 2018).

- Understandability denotes the ability to make a user understand AI systems function without requiring technical or internal knowledge of the system (Arrieta et al., 2020).
- Transparency refers to the algorithm's (or system's) technical intelligibility (Lepri et al., 2018; Gerlings et al., 2021). A model is considered transparent if it can explain how the system works even if it does something unexpected (Arrieta et al., 2020; Vilone and Longo, 2021). A transparent system is, therefore, the exact opposite of a "black-box" model.
- Interpretability describes the extent to which the system is intelligible to the user (Miller, 2019). In this context, interpretability refers to more than one concept, e.g., mathematical approaches like approximations, visualizations, natural language, explanations by examples, etc., to achieve this goal (Lipton, 2018).
- Comprehensibility refers to the ability of an algorithm to semantically and structurally represent an AI's recommendations as a human expert would (Arrieta et al., 2020).

Moreover, the distinction between trust and reliance concerning XAI is often neglected. This leads to complications, as several studies often discuss trust but measure and understand the concept in entirely different ways (Schmidt et al., 2020; Yin et al., 2019).

- Trust is defined as an attitude and measured subjectively (e.g., by interviews or Likert scale questionnaires) (Scharowski et al., 2022; Mohseni et al., 2021).
- Reliance, in contrast, reflects human actual behavior and is objectively observable and measurable (e.g., by the percentage of users' consent to AI advice) (Scharowski et al., 2022).

It is essential to distinguish this because research recognizes a discrepancy between attitude and behavior, indicating that changes in mentality do not always correspond with changes in behavior and vice versa (Ajzen, 2001; Kohn et al., 2021).

However, explanations serve as the key approach to improving the transparency of interaction between humans and AI systems (Miller, 2019; Wang and Yin, 2021). In theory, an explanation should explain the reasons for the decision to the recipient to convince or help the recipients better assess their decision to follow or reject the given advice (Gregor and Benbasat, 1999; Bansal et al., 2021). In practice, the impact of AI explanations on decision-making is mixed. Although some studies highlight the benefits of explanations—such as improving transparency, decision quality, and trust—others reveal potential drawbacks (Dzindolet et al., 2003; Ehrlich et al., 2011; Ribeiro et al., 2018; Lai and Tan, 2019; Zhang et al., 2020; Carton et al., 2020; Chu et al., 2020; van der Waa et al., 2021). These negative consequences are often denoted as under- or overreliance. Underreliance occurs when users fail to grasp an explanation, for instance, performing below the numerical certainties of an AI recommendation (Carton et al., 2020; Zhang et al., 2020), or when algorithm aversion adversely impacts decision-making (Dietvorst et al., 2018; Jussupow et al., 2020). On the other hand, an explanation can mislead users into believing in the general accuracy of AI advice, increasing the reliance even on an erroneous AI recommendation (Ehrlich et al., 2011; Vilone and Longo, 2021; Rebholz et al., 2024), a pattern called overreliance. In conclusion, it is still uncertain whether AI explanations help users make correct decisions or contribute to under- or overreliance on AI advice (Mohseni et al., 2021; Schemmer et al., 2022). These statements confirm that explanations are not employed in the practical environment to the extent dictated by theoretical expectations. Humans must understand and evaluate the underlying rationality of a recommendation to appropriately calibrate their trust and reliance (Wang and Benbasat, 2016). These considerations reveal a fundamental research gap regarding the actual design of explanations.

Compared to humans, AI systems also struggle to communicate decision rationales, especially in complex models, which prompted efforts in XAI to improve transparency in deep learning (Burrell, 2016; Liao et al., 2020; Baird and Maruping, 2021). Furthermore, a simple technical solution is not sufficient to ensure optimal decision-making, as humans do not always behave reasonably (Zedelius et al., 2017; Abdul et al., 2018; Bonnefon and Rahwan, 2020; Jussupow et al., 2020; Buçinca et al., 2021; Fügener et al., 2022). Therefore, improving a system's accuracy does not inherently result in better human-AI collaboration (Dzindolet et al., 2003; Wilder et al., 2020; van der Waa et al., 2021). This challenge is further intensified by the fact that, as AI proliferates across multiple domains, it is no longer only AI specialists who engage with these systems—beginners and non-expert users are increasingly required to work with AI as well (Laupichler et al., 2022; Ng et al., 2021a). Thus, humans often find themselves unavoidably entering increasingly close interactions with AI systems (Lu and Zhang, 2024). Problems in non-expert user interaction with AI arise due to this technological invasion. For example, AI explanations are often not tailored for non-expert users (Abdul et al., 2018; Miller, 2019), despite evidence that the user's expertise influences the requirements regarding the scope and nature of the explanation (Gregor and Benbasat, 1999; Liao et al., 2020; Schoeffer et al., 2024). Nonetheless, the interactions between non-expert users and AI are understudied and require further research (Long and Magerko, 2020; Ng et al., 2021a).

Although there is consensus on the necessity and purpose of XAI, considerable ambiguity remains regarding the practical design of explanations and non-expert users' perceptions of these explanations. No single method for XAI has yet emerged, and research continues to explore ways to improve XAI (Pedreschi et al., 2019). Building on anthropomorphism in the human-AI context, which has often been overlooked in the XAI literature, it is plausible to apply human-to-human explanation techniques as a promising starting point. Considering the explanatory methods of human advisors, one can conclude that humans tend to express predictions using numbers or words, affecting perception (Mislavsky and Gaertig, 2022). Applying this to XAI, an ongoing debate exists on whether an AI should provide numerical or verbal explanations. Several experts insist that AI should explain its advice like a human would (Weiner, 1980; De Graaf and Malle, 2017; Byrne, 2019; Miller, 2019). The key point is that verbal AI explanations are easier for humans to grasp, helping them form accurate mental models of the system and adequately adjust their reliance on the advice (De Graaf and Malle, 2017). However, even during human-to-human interaction, a preferential paradox arises as individuals feel more comfortable expressing uncertainties in verbal form when taking the role of recommender, whereas humans tend to prefer numerical certainties of a recommender to make a decision (Erev and Cohen, 1990; Wallsten, Budescu, Zwick, and Kemp, 1993). Scholars supporting this argument highlight that individuals frequently find it challenging to accurately interpret verbal descriptions of uncertainties and probabilities, advocating for using numerical statements instead (Budescu and Wallsten, 1985; Wallsten, Budescu, and Zwick, 1993; Wintle et al., 2019). Therefore, the following research question arises:

RQ1: Are humans more inclined to understand and process verbal versus numerical explanations of an AI system, and will the type of explanation affect their appropriate reliance?

While AI-generated explanations are intended to enhance transparency and improve decision-making, users often struggle to assess them critically (Lai and Tan, 2019;

Bansal et al., 2021). Since cognitive effort usually feels unpleasant to humans (David et al., 2024), users tend not to engage rationally and cognitively with AI explanations (Buçinca et al., 2021). As algorithms are susceptible to biases and often do not detect guidance errors, users must think analytically and critically about explanations to prevent misjudgments (Rebholz et al., 2024). Furthermore, humans frequently misinterpret an explanation as a marker of expertise and over-rely on flawed AI advice (Bansal et al., 2021; Lai and Tan, 2019). Thus, despite the general improvement in performance when users are provided with explanations of an AI system (van der Waa et al., 2021; Zhang et al., 2020), recent research presents mixed findings on whether such explanations help users identify incorrect recommendations or rather make both correct and incorrect recommendations seem more convincing (Ehrlich et al., 2011; Vilone and Longo, 2021). In particular, non-expert users often struggle to critically review AI recommendations due to a lack of expertise (Gregor and Benbasat, 1999; Guidotti et al., 2018). Hence, the question of when it is appropriate to follow the system's advice remains (Ehrlich et al., 2011). To avoid errors, users must consider the explanation thoughtfully, which requires cognitive effort and time (Bunt et al., 2012; Buçinca et al., 2021; Gajos and Mamykina, 2022). Therefore, humans need to apply analytical thinking (slow-thinking) instead of relying on their gut feeling (fast-thinking) (Kahneman, 2011). Since users seem to strategically navigate the trade-off between the benefit and cost ratio of the mental effort required to validate the explanation (Bunt et al., 2012; Vasconcelos et al., 2022), the active demand for an explanation may enhance their willingness to do so (Buçinca et al., 2021). Therefore, an actively requested explanation might be valued more and scrutinized thoroughly. In other words, this shift in control could modify interaction dynamics by allowing users to manage incoming information and decide when they need further clarification through explanation. This logic highlights the following research question:

RQ2: How does the accessibility of an AI explanation - whether immediately displayed or only available on demand - affect the decision-making process of non-expert users?

As AI becomes more sophisticated, concerns about its control, fairness, and unintended consequences lead to growing skepticism toward the technology (Ågerfalk, 2020). This has resulted in AI aversion—a reluctance to accept AI-driven recommendations due to perceived risks and biases (Castelo et al., 2019; Yeomans et al., 2019). Consequently, negative attitudes toward AI often diminish its potential benefits (Dietvorst et al., 2015; Dietvorst et al., 2018; Jussupow et al., 2020). Drawing on Kahnemann's dual system theory of reasoning, it proposes different conditions that might arouse humans' active information processing and analytical thinking to prevent them from relying solely on their feelings (Kahneman, 2011). Besides the availability of an AI-generated explanation, the response time necessary for the AI to complete a task may affect the user's cognitive approach and alter human-AI collaboration dynamics. Humans attribute different ways of thinking to different task domains (Bonnefon and Rahwan, 2020). Therefore, humans are more likely to reject AI recommendations in task domains that they believe require intuition or gut feeling, such as hiring (subjective tasks), and are more likely to follow algorithms within a technical or mathematical task domain (objective tasks) (Lee, 2018; Castelo et al., 2019). Despite existing research that indicates that AI systems surpass human experts even in subjective tasks, there remains a tendency for humans to dismiss AI suggestions in these cases (Yeomans et al., 2019). This suggests that humans transfer the expected skill set and way of thinking for the task to intelligent machines (Bonnefon and Rahwan, 2020; Booch et al., 2021; Rossi and Loreggia, 2019). So, humans assume that they are primarily capable of performing subjective tasks, and objective tasks are associated with higher cognitive effort, whereas this assumption is reversed in the case of intelligent machines. An initial study shows that when algorithms exhibit extended response times for objective tasks, humans often perceive this as a malfunction since machines are expected to perform these tasks efficiently (Efendić et al., 2020). Thus, analyzing the impact of AI response time on human-AI interaction in both fast- and slow-thinking tasks could offer a more nuanced understanding of mitigating algorithm aversion, particularly in subjective tasks. Which in turn raises the following research question:

RQ3: Which effect does the AI response time have on algorithm aversion for slow-thinking and fast-thinking tasks?

Until now, most research on human-AI interaction has primarily emphasized performance metrics (Fügener et al., 2021; Millet et al., 2023; Chen and Chan, 2024; Bohren et al., 2024) and dyadic interactions (Seeber et al., 2020; O'Neill et al., 2022; Bouschery et al., 2023; Boussioux et al., 2024; Doshi and Hauser, 2024), neglecting the processes that lead to those outcomes and that tasks occur frequently in group settings. This focus overlooks how AI might influence team dynamics, particularly in collaborative contexts. Although in traditional research on human-only teams, the input-process-output (I-P-O) model was developed to investigate the essential function of team processes in converting inputs into outputs, the process element is rarely considered in human-AI interaction (O'Neill et al., 2023), despite the circumstance that team processes are crucial for successful collaboration (Sjøvold et al., 2022; Marks et al., 2001; Mathieu et al., 2020). However, it remains unclear whether individual and team creativity is enhanced (e.g., idea sharing, integrating perspectives, cognitive stimulation) (Pinsonneault et al., 1999; Dugosh et al., 2000; Paulus, 2000) or hindered (e.g., collaborative fixation, production blocking) (Diehl and Stroebe, 1987; Kohn and Smith, 2011) by social interactions (Amabile, 2018). Therefore, this gap is especially critical in the creative task domain, where the process is a central component of innovation and novel value creation. As generative AI (GenAI) continues to evolve and expand its capabilities across domains such as text, imagery, sound, and video (Millet et al., 2023), its integration into creative workflows is becoming increasingly common and influential. For instance, large language models (LLM) now serve as effective collaborators in tasks such as academic and narrative writing (Wan et al., 2024), helping users generate novel ideas and even new knowledge (Raisch and Fomina, 2025). Hence, GenAI facilitates new forms of human-AI cooperation in open-ended, non-routine, and highly creative tasks (Chen and Chan, 2024), and in some cases, even outperforms humans in terms of idea originality or diversity (Bohren et al., 2024). Nevertheless, human evaluators often respond adversely when they learn that an AI system contributed to creative output, devaluing the work due to persistent associations between creativity and human effort (Demir et al., 2024; Bohren et al., 2024; Millet et al., 2023). Furthermore, AI tends to be more effective when it complements human creativity than when working independently (Anantrasirichai and Bull, 2022). In summary, GenAI is likely to radically change the way creators conceive and produce ideas (Epstein et al., 2023). However, the current literature lacks information on how collaboration with AI in creative tasks might impact the creative process between multiple humans, either by inhibiting or facilitating it (Amabile, 2020). As a result, a simple but fundamental research question remains unconsidered:

RQ4: How does AI influence team dynamics during creative tasks?

1.2. Presentation of Papers

This dissertation addresses the four research questions through four separate research studies, each presented in a dedicated chapter. All manuscripts vary in length, scope, and style due to their preparation for different academic publications. Each study's research objectives and importance are addressed by summarizing the methodology, key findings, and contributions. Following the first chapter, this dissertation is structured along the following papers:

1. J. Papenkordt, A. C. Ngonga Ngomo, K. Thommes (2025) **"Are numerical or verbal explanations of AI the key to appropriate user reliance and error detection? An experimental study with a classification algorithm"**.

   Faced with AI's significant advantages and drawbacks, there's been a growing demand for making AI systems explainable (Adadi and Berrada, 2018; Thiebes et al., 2021). Situations in which humans are forced to rely on AI blindly due to its complexity must be avoided. An explanation must act as a bridge between the user and the AI system, accurately reflecting the decision while simultaneously being comprehensible to the user (Guidotti et al., 2018). Achieving XAI in practice remains a challenge (Vilone and Longo, 2021; Guidotti et al., 2018). First, due to bounded human rationality, users struggle to process complex information, requiring a trade-off between accuracy and interpretability in AI models (Gerlings et al., 2021). Second, although XAI frameworks can technically explain outputs, these explanations often remain incomprehensible to non-experts, as most systems do not tailor explanations to user's levels of expertise (Guidotti et al., 2018; Gregor and Benbasat, 1999; Schoeffer et al., 2024; Liao et al., 2020). Third, while explanations may improve user understanding, their impact on appropriate reliance remains ambiguous—particularly as explanations can enhance overreliance on flawed recommendations (Miller, 2019; Rebholz et al., 2024; Vilone and Longo, 2021). Consequently, explanations often fall short of their anticipated expectations, indicating an underappreciated potential for improvement (Gregor and Benbasat, 1999).

   Therefore, this research investigates how different forms of AI-generated explanations—precisely, verbal explanations (e.g., natural language justifications) and numerical explanations (e.g., certainty values), or a combination of both—affect the behavior of non-expert users, particularly in terms of reliance on AI recommendations, decision accuracy, and the ability to detect AI errors (**RQ1**). Drawing on theories from human decision-making (such as the law of less work), social cognition (such as anthropomorphism), and XAI, the paper addresses the pressing need to balance comprehensibility and transparency in AI explanations, particularly for non-expert users.

Concretely, the research employs a between-subjects experimental design, with 441 participants completing ten classification tasks, either with the support of a self-developed AI offering varying forms of explanation or without any AI support. Participants were randomly assigned to one of five treatment groups. The five experimental conditions include (1) verbal explanation, (2) numerical explanation, (3) both verbal and numerical, (4) a control group with AI advice but no explanation, and (5) a baseline group without any AI support. Task complexity was deliberately high, and an intentional AI error was introduced in task 4 to observe error detection and the recovery of the reliance. By doing so, this study leads to the following key findings:

- AI support significantly increases decision accuracy compared to the baseline group, regardless of the explanation format.
- Users often under-relied on AI when it was correct and over-relied when it was wrong, reflecting biases in human-AI interaction.
- Verbal explanations lead to higher user reliance and more correct decisions overall, suggesting they are easier for non-experts to process. However, this comes at the cost of increased over-reliance.
- In contrast, numerical explanations tend to be more effective in detecting AI errors. On the one hand, numerical indicators sensitize users to AI uncertainty; on the other hand, reliance on numerical explanations does not proportionally reflect the certainty provided, highlighting gaps in the statistical intuition of users.
- When verbal and numerical explanations were presented together, user reliance increased even further—regardless of whether the recommendation was correct or incorrect.

The study contributes to the literature by empirically demonstrating the trade-offs between explanation types regarding reliance and error detection. This provides deeper insights into human cognitive patterns when engaging with AI. It cautions against one-size-fits-all XAI approaches and underscores the need for explanation designs that account for user expertise and cognitive constraints. This research demonstrates that reliance is nuanced and varies depending on the type of explanation provided. These findings also illustrate that users frequently act irrationally in AI interactions, exhibiting underreliance when the AI is accurate and overreliance by not engaging cognitively and critically with AI explanations. In summary, this investigation offers actionable insights for designing human-centered AI systems that support appropriate reliance, especially in low-stakes, high-frequency decision contexts involving non-experts.

2. J. Papenkordt (2024) **"Navigating Transparency: The Influence of On-demand Explanations on Non-expert User Interaction with AI"**.

   Inspired by the findings of the first paper, which demonstrated that explanation format significantly affects user reliance and error detection, this second paper examines a crucial but underexplored design decision in XAI: whether the accessibility of an AI explanation affects the decision-making process of non-expert users (**RQ2**). While explanations often support decision-making, recent research highlights that users frequently either neglect them due to negative feelings toward AI (Dietvorst et al., 2018; Jussupow et al., 2020) or over-rely on them by failing to critically evaluate them (Gajos and Mamykina, 2022; Buçinca et al., 2021; Bunt et al., 2012). In particular, verbal explanations, which are becoming increasingly common due to the advances of LLM, are relatively intuitive but can lead users to accept recommendations without sufficient scrutiny (Lebedeva et al., 2023; Miller, 2019; Rebholz et al., 2024). Despite growing attention to human-centered AI, most XAI systems still present explanations by default without considering that this automatic transparency may induce information overload (Bawden and Robinson, 2009), reactance of contrary advice (Fitzsimons and Lehmann, 2004), or blind reliance as users perceive the explanation as a general signal of quality (Bansal et al., 2021; Lai and Tan, 2019). In particular, non-expert users might be more sensitive to these issues as they lack the expertise to critically scrutinize the AI recommendation directly (Gregor and Benbasat, 1999; Guidotti et al., 2018; Ng et al., 2021a). Therefore, this study proposes that the effectiveness of explanations may depend not only on how an explanation is designed but also on when and how users choose to access those explanations. Drawing on dual system theory and the cost-benefit principle, one might argue that the active request for an explanation enhances the user's willingness to engage with the explanation cognitively.

The study design builds directly on the setting and data from the first paper by keeping all experimental conditions constant and expanding the study by two experimental groups. 151 participants were randomly assigned to one of the two new treatments. The task design mirrored the first study: high task complexity, monetary incentives, task order, and an intentional AI error in task 4. The subjects completed the same ten AI-supported classification tasks. The novel treatments encompass (6) the option to request a verbal explanation of AI advice and (7) the immediate presentation of numerical certainty for AI advice, alongside the on-demand option for a verbal explanation. By conducting different analyses, the study provides the following main results:

- Explanations are underutilized, even when they would be beneficial. About 50% of the participants in each group did not demand a single verbal explanation during all tasks.
- Personal factors (such as educational level, age, or attitude toward AI) significantly affect the demand for verbal explanations.
- Contextual factors (such as the perceived difficulty of the task or the certainty of the AI) influence the extent to which verbal explanations are demanded, but certainly not to the expected degree.
- Participants who requested explanations showed a higher reliance on AI recommendations, similar to those who received explanations automatically.
- By significantly increasing the decision times of the user, on-demand explanations appear to promote analytical thinking (slow-thinking). However, this greater cognitive effort does not consistently improve error detection.

This research expands the literature by empirically examining the role of explanation accessibility, a previously underexplored but theoretically plausible factor, in shaping the human-AI interaction. By shifting control to the user through an on-demand explanation option, the study uncovers how personal characteristics and contextual factors influence the demand and use of an explanation. The fact that non-expert users often abstain from requesting explanations despite clear performance benefits suggests that users may rely on decision heuristics instead of applying strategic thinking. Although the on-demand option significantly increases longer decision times (a proxy for analytical thinking) as a practical design feature, the findings highlight that users often struggle to optimize the cost-benefit trade-off and make the optimal strategic decision.

3. A. Lebedeva, J. Kornowicz, O. Lammert, J. Papenkordt (2023) **"The Role of Response Time for Algorithm Aversion in Fast & Slow Thinking Tasks"**.

   The study builds on dual system theory, distinguishing between fast- and slow-thinking (Kahneman, 2011). While this theory has traditionally been applied to understand human decision-making, Bonnefon and Rahwan (2020) and Rahwan et al. (2019) argue that it can be extended to human-machine interaction. Specifically, they propose that humans anthropomorphize AI agents by attributing to them similar modes of fast- or slow-thinking and make trust-related inferences based on perceived cognitive processes (Bonnefon and Rahwan, 2020). Although algorithms do not "think" (Booch et al., 2021), different scholars suggest that the response time may serve as a salient cue in this process, tempting humans to assess the invested ("cognitive") effort of the algorithm (Bonnefon and Rahwan, 2020; Efendić et al., 2020; Park et al., 2019). A fast AI response might be considered effortless, whereas a longer response time may signal deliberate, effortful reasoning. This attribution may affect how individuals judge the suitability of AI for a particular task domain. If an AI appears to work in a way that matches the expected demands of the tasks, users may be more inclined to rely on recommendations. This theoretical proposition might expand prior research on algorithm aversion and diminish aversion toward AI based on the task domain. Different studies reveal that humans are more likely to rely on algorithms in tasks perceived as objective, mechanical, or analytical, and less likely in tasks perceived as subjective or intuition-based (Castelo et al., 2019; Lee, 2018), although AI already excels at subjective tasks domains (Yeomans et al., 2019). By linking the theoretical assumptions about the mapping of dual system theory to the human-AI context and the practical findings on algorithm aversion, this research investigates how the response time of an AI might reduce algorithm aversion, especially in subjective task domains (**RQ3**). For instance, we propose that in logically approached tasks (objective task domain), humans anticipate a rapid AI response, as these tasks are considered trivial for algorithms.

We conducted a 2x2 between-subjects laboratory experiment with 116 students to investigate these assumptions. The experiment manipulated the AI response time (short vs. long) and task type (fast-thinking vs. slow-thinking) in a controlled advice-taking setting based on the Judge-Advisor-System framework (Bonaccio and Dalal, 2006). During the experiment, participants had to solve nine estimation tasks from three different domains. The tasks were drawn equally from three domains: estimating the number of lentils in a glass (Park et al., 2019), the weight of football players (Gino and Moore, 2007), and the distances between cities (Hofheinz et al., 2017). To encourage intuitive judgments in the fast-thinking treatment, the tasks provided only minimal visual cues, while in the slow-thinking treatment, the task included quantitative information to promote analytical reasoning. Algorithm aversion was measured using the advice-taking index, which captures how participants adjusted their initial estimates toward the AI's recommendation (Bailey et al., 2022; Hofheinz et al., 2017). The precision of the final estimation was monetarily incentivized, and additional variables (such as perceived recommendation quality (Gino et al., 2012)) were controlled. The comparison of different groups provides the following results:

- Contrary to our assumptions, longer AI response times reduce algorithm aversion across both fast- and slow-thinking tasks except in one of the three domains.
- The task domain matters, but only in fast-thinking tasks. Within these, advice-taking differed significantly between all domains except in one instance. No such variation was found in slow-thinking tasks, suggesting that domain effects are muted when users receive similar information and may start to think analytically.
- Interestingly, the participants felt less confident in the "Lentils" task domain, where the effect of longer response time on the advice-taking index was strongest, suggesting that lower confidence in one's capabilities amplifies reliance on seemingly effortful AI advice.

This research broadens the literature on algorithm aversion in several ways. First, it connects the dual system theory with human-AI interaction and empirically investigates how human thought patterns might influence human-AI collaboration. In particular, it demonstrates how the response time of an AI system can serve as a meaningful signal, influencing user acceptance of AI advice-taking. Interestingly, the results also highlight the necessity of future research, as they contradict initial expectations rooted in the literature and human-to-human interaction, suggesting that longer response times increase reliance on AI advice, even in slow-thinking tasks. Moreover, this paper offers practical implications for AI system design, implying that calibrated response delays might reduce algorithm aversion and improve decision-support outcomes. Lastly, our research suggests that domain-specific effects may be mitigated if users are encouraged to think analytically, for example, by presenting additional information.

4. J. Papenkordt, J. Dahlke, N. Neef, S. Zabel (2025) **"Exploring the impact of AI on team collaboration dynamics in creative decision-making"**.

GenAI has emerged as a transformative technology, producing outputs that are nearly indistinguishable from human-authored work (Giray, 2024; Xu and Sheng, 2024; Doshi and Hauser, 2024; Zercher, Jussupow, and Heinzl, 2025; Dell'Acqua et al., 2025). This leads to human-AI collaboration evolving; for example, humans often no longer perceive AI as just a tool but as a teammate (Seeber et al., 2020; De Freitas et al., 2025; Dell'Acqua et al., 2025; Zercher, Jussupow, and Heinzl, 2025) or work interactively with an AI to perform knowledge-intensive tasks that go beyond routine automation (Köbis and Mossink, 2021; Epstein et al., 2023; Wan et al., 2024; Doshi and Hauser, 2024). Early studies have focused on the impact of GenAI on dyadic human-AI teams (Bouschery et al., 2023; Boussioux et al., 2024; Doshi and Hauser, 2024; O'Neill et al., 2022; Seeber et al., 2020), exploring predominantly the effect of input variables (such as task characteristics or user experiences) (Walliser et al., 2019; McNeese et al., 2018; Wright et al., 2018) or outcome variables (such as performance or time spent) (Dell'Acqua et al., 2025; Doshi and Hauser, 2024; Zhou and Lee, 2024). To date, the mediating mechanisms that transform input into output, as well as the team dynamics that unfold between multiple humans during collaboration with AI, remain underexplored. In this context, creative tasks may pose specific challenges, as previous studies have shown ambivalent results regarding the impact of AI support on creative performance (Doshi and Hauser, 2024; Zhou and Lee, 2024; Fügener et al., 2021). Moreover, earlier systems mainly facilitated divergent thinking by boosting individual or team creativity through new stimuli (Wang and Nickerson, 2017), GenAI can now shape convergent thinking by directly affecting creative outputs (Epstein et al., 2023). Therefore, this research investigates how GenAI influences team dynamics during creative tasks (**RQ4**).

To address our research question, we conducted a controlled laboratory experiment using GenAI. Specifically, we utilize the modern large-scale generative model ChatGPT4o to support treatment groups in creating a three-act short story. The participants were randomly divided into 30 teams of four (N = 120), with equal proportions of students drawn from the universities of Twente, Hohenheim, and Paderborn, and distributed evenly across the following three conditions by random selection: (1) no AI assistance (control) during the task, (2) full AI assistance throughout the task, and (3) partial AI assistance limited to the second act of the task. By contrasting collaboration settings with and without AI within the partial AI treatment and between the control and full AI treatment, we can delve deeper into how AI impacts team collaboration dynamics both temporarily and permanently. To further ensure comparability, each session was audio- and video-recorded, each act was limited to 20 minutes, and all teams began with the same introduction. Moreover, to minimize the potential for experimenter bias and expectancy effects, the student assistants conducting the experimental sessions were not informed of the specific research objectives. Through various analyses of the team collaboration dynamics, the study yields these exploratory results:

- Permanent AI assistance significantly alters team processes, marked by a reduction in problem-focused statements and an increase in socio-emotional ones, suggesting a shift in cognitive effort and interpersonal team dynamics.
- Permanent AI support tends to reshape communication structure and patterns, with less bilateral (dyadic) exchange and more group-directed communication, implying a more balanced participation and a redistribution of conversational influence toward the team or AI as a whole.

- Temporary AI assistance appears to influence conversational content by reducing problem-focused statements and increasing procedural ones, indicating a short-term renegotiation of the workflow without affecting underlying communication structure and turn-taking patterns.
- Creativity ratings showed no significant variation across treatment conditions, indicating AI collaboration, whether permanent or temporary, neither hindered nor enhanced team creativity. However, permanent AI had a positive impact on the content, structure, and expression of the story.

Overall, this research contributes to the literature on human-AI teaming by offering one of the first empirical investigations into how GenAI transforms team collaboration dynamics in creative tasks. By doing so, it provides initial insights into the collaboration between multiple humans and AI by investigating how team dynamics, in terms of conversational content, structure, and turn-taking patterns, change during a creative task setting. The influence of the permanent AI integration suggests a paradigm shift in team dynamics from hierarchical, dyadic exchanges to flatter, group-oriented interaction patterns, inviting a reconceptualization of traditional constructs such as leadership, dominance, and communication centrality in AI-supported teams. Furthermore, the research provides initial insights into the impact of temporal continuity in AI integration, as permanent AI support gradually reshapes structural and behavioral dynamics. Temporary collaboration with AI appears insufficient to lead to lasting change, highlighting the need for more long-term experimental research on this topic. From a practical perspective, the findings carry initial implications for organizations seeking to harness AI in collaborative work. Rather than viewing AI as an external add-on or isolated decision aid, this research suggests that AI becomes entangled in the team dynamics, affecting not only what is said but who speaks, how turn-taking unfolds, and how roles and responsibilities are distributed. Regarding the similar creative outputs of the teams, it appears that while AI support tends to affect team collaboration dynamics to varying degrees, it does not undermine the underlying creative workflow of the teams. However, further research is required, as the creative storytelling task likely reflects the inherent subjectivity in evaluating creative work, and therefore makes it challenging to assess creative performance objectively, as the low consistency between our ratings indicates.

1.3. Contributions to Research and Practice

In an era where AI technologies are becoming deeply embedded in individual, team, and organizational decision-making processes (Thiebes et al., 2021; Poursabzi-Sangdeh et al., 2021; Raisch and Fomina, 2025), this dissertation contributes to a broader understanding of how humans interact with intelligent systems in diverse contexts and modalities. The complexity, opaqueness, and fallibility of AI systems, particularly in high-stakes scenarios, demand that users, regardless of their expertise, be empowered to interact with these technologies in a thoughtful manner (Burrell, 2016; Laupichler et al., 2022; Gerlings et al., 2021; Mohseni et al., 2021). Yet the rapid pace of AI development risks outpacing both societal readiness and individual cognitive capacity (Fügener et al., 2021; Fügener et al., 2022). Therefore, successful implementation hinges on more than technical excellence; it requires a nuanced understanding of how humans interact with, rely on, and co-create with AI systems (Gunning and Aha, 2019; Arrieta et al., 2020; O'Neill et al., 2023). Across four studies, this dissertation offers a multi-level perspective on how individuals (*Paper 1*, *Paper 2*, *Paper 3*) and teams (*Paper 4*) interact with AI. Therefore, this thesis integrates perspectives from behavioral decision-making, XAI, dual system theory, and team science to examine how individuals cognitively and socially engage with AI systems.

Although XAI has been positioned as a promising response to the opaqueness of AI systems, this line of thinking repeatedly neglects the fact that users do not engage with explanations rationally and reflectively. By applying social and psychological frameworks to the IS literature, this thesis attempts to acknowledge that humans frequently rely on intuition, emotion, and heuristic reasoning (Kahneman, 2011; Zedelius et al., 2017; Bonnefon and Rahwan, 2020; Kahneman et al., 2021; Buçinca et al., 2021). The dissertation reconsiders explanation design as a cognitive and social interface, rather than a mere technical aspect, influencing reliance, cognitive engagement, and user behavior in human-AI interaction. Therefore, it contributes to the literature by empirically examining the type of explanations (*Paper 1*), the accessibility of explanations (*Paper 2*), and the timing of explanations (*Paper 3*) in the human-AI context. Especially, by focusing on non-expert user interactions with AI, this research addresses an underexplored but increasingly crucial field in research and practice, as increasing utilization of AI inevitably leads to non-expert users engaging with it (Miller, 2019; Long and Magerko, 2020; Ng et al., 2021a; Lu and Zhang, 2024).

Since user expertise influences the requirements of an explanation (Gregor and Benbasat, 1999; Liao et al., 2020) and humans frequently anthropomorphize AI (Mislavsky and Gaertig, 2022; Seeber et al., 2020), *Paper 1* investigates behavioral effects of two common (verbal and numerical) forms of human-to-human interaction in human-AI interaction. The findings demonstrate that the types of explanation matter because they result in distinct behavioral patterns and appear to reflect well-established differences in mental processing. These observations contribute to the ongoing debate on whether an AI explanation should be verbal or numerical (Miller, 2019; De Graaf and Malle, 2017; Gneiting and Katzfuss, 2014; Byrne, 2019; Wintle et al., 2019; Mislavsky and Gaertig, 2022). Consistent with prior work highlighting the persuasive power of verbal communication (Rebholz et al., 2024; Miller, 2019), the findings reveal that verbal explanations increase reliance on AI recommendations, even when they are incorrect. Numerical explanations, while more precise, are often underutilized by non-expert users, who may lack the statistical knowledge to interpret uncertainty correctly (Gigerenzer et al., 2005; Fu et al., 2022; Pedreschi et al., 2019; Liao et al., 2020). In line with previous research, these findings highlight that technical improvements to the intelligent system alone are insufficient to enhance human-AI collaboration (Wilder et al., 2020; van der Waa et al., 2021). If explanations are frequently treated as general indicators of quality (Vilone and Longo, 2021; Bansal et al., 2021), they must be used and designed with caution, as increased transparency can intuitively lead to blind reliance. This is especially concerning since LLMs produce fluent justifications that may discourage critical scrutiny and foster a misleading impression of comprehension (Rebholz et al., 2024; Spitzer et al., 2024). As explanation design is not just a matter of information delivery, researchers and practitioners must investigate the psychological and contextual factors that may drive or hinder users' cognitive engagement with explanations.

This dissertation presents a first attempt to address this issue by modifying the interaction dynamics, enabling users to control the amount of incoming information and decide when they need or require additional clarification through an explanation (*Paper 2*). Shifting the control to demand an explanation to the user was assumed to enhance the user's willingness to engage cognitively with the explanation by increasing the user's autonomy (Buçinca et al., 2021). However, many non-expert users underuse this functionality, even when explanations would have significantly improved performance. Non-expert users seem to struggle with the cost-benefit trade-off of engaging with explanations. This behavior appears to be shaped by a combination of contextual factors (e.g., perceived task complexity) and individual characteristics (e.g., educational background, attitude toward AI), aligning with the research on cognitive unpleasantness and mental economy in the interaction of AI (David et al., 2024; Bunt et al., 2012). From an organizational perspective, this underscores the importance of promoting AI literacy in the workforce (Long and Magerko, 2020; Laupichler et al., 2022; Ng et al., 2021b). On a broader scale, the findings, in combination with other mixed results on the effectiveness of explanations (Mohseni et al., 2021; Schemmer et al., 2022; Schoeffer et al., 2024), raise the question of whether an explanation is always necessary, given that users tend not to thoroughly review them. Nonetheless, this seems to be scarcely feasible in practice because users may start to question the system's overall advantage if they need to evaluate each recommendation critically. Therefore, a promising approach for future researchers and practitioners could be to design adaptive, context-aware explanation systems where the display of an explanation is triggered by high AI uncertainty or a high-stakes decision (Liel and Zalmanson, 2023). Additionally, users may be nudged to think analytically and critically through presented social comparisons with previous users (Allcott and Rogers, 2014; Allcott and Kessler, 2019; Bicchieri and Dimant, 2022), the provision with alternative AI recommendations (Miller, 2023) or the application of gamification elements (Zhan et al., 2022).

Extending this line of research, this dissertation investigates how subtle interface-level features might affect the thought patterns of users and consequently human-AI collaboration (*Paper 3*). Therefore, this research contributes to the literature by combining existing results on task types (Lee, 2018; Castelo et al., 2019; Yeomans et al., 2019) and response times (Efendić et al., 2020; Park et al., 2019) with the theoretical consideration proposed by Bonnefon and Rahwan (2020) in applying the dual system theory (Kahneman, 2011) to human-AI interaction. Contrary to theoretical grounded expectations and comparable studies (Efendić et al., 2020), the results suggest that longer AI reaction times reduce algorithm aversion, regardless of the planned perception of the task. From a research perspective, this calls for further studies in various tasks and experimental settings. From a practical standpoint, these findings suggest that response time is a subtle yet powerful design lever to mitigate algorithm aversion (Dietvorst et al., 2018; Jussupow et al., 2020). Moreover, domain-specific biases (Castelo et al., 2019; Lee, 2018; Mahmud et al., 2022) could potentially be diminished by providing additional information. Furthermore, this research suggests that deliberate delays may boost user acceptance in high-complexity or low-confidence scenarios. This insight opens new pathways for behaviorally informed interface design, even without improving the underlying algorithm. In practice, explanations may not need to be altered in content; instead, adjustments in timing or tone could be used strategically. For instance, future research should investigate how delays in response time should be communicated (e.g., visually, verbally, or numerically) and whether they might backfire in time-sensitive or repetitive contexts. Regarding the advances in GenAI and voice assistant systems (Poushneh, 2021), it is conceivable that an AI that communicates its "thinking procedure" in a human-like manner could enhance trust by reinforcing anthropomorphic perceptions (Horstmann et al., 2018).

Building on this progress in AI, an upcoming research stream conceptualizes AI as an implicit team member embedded in social and collaborative processes (De Freitas et al., 2025; Dell'Acqua et al., 2025; Zercher, Jussupow, and Heinzl, 2025). To contribute to the new strand of literature, *Paper 4* highlights the interaction dynamics that emerge when GenAI is embedded in multi-human teams. Moving beyond typical dyadic (one-human-one-AI) setups (Bouschery et al., 2023; Boussioux et al., 2024; Doshi and Hauser, 2024; Seeber et al., 2020), this thesis extends the research by focusing on team dynamics that transform input into output in creative decision-making, which have not yet been adequately explored (O'Neill et al., 2022; O'Neill et al., 2023). The results of this paper suggest that the presence of AI subtly but systematically alters how teams collaborate by influencing the content, structure, and flow of the interaction. For example, permanent AI integration seems to be associated with shifts in communication content (e.g., by altering the cognitive division of labor or enhancing socioemotional statements), a redistribution of influence from dominant individuals to the team, and a move from dyadic exchanges to more team-oriented forms of interaction. These patterns, in combination with comparable initial studies regarding team climate (Zercher, Jussupow, Benke, et al., 2025) and team sociality (Dell'Acqua et al., 2025) signal a structural transformation in team dynamics that may have implications for how leadership, coordination, and role differentiation evolve. Overall, *Paper 4* represents an initial step toward theorizing the human-AI interaction at the team level. Future research should expand on these findings by exploring other team configurations, task domains, and longitudinal effects of sustained AI collaboration. Applying these insights from a practical perspective might induce the need to rethink conventional teamwork models in creative and knowledge-intensive contexts. As AI increasingly shapes not only individual productivity but also collective behavior, organizations must consider how, when, and to what extent AI should be embedded in collaborative workflows.

1.4. Limitations

While this dissertation offers novel insights into human-AI interaction at both individual and team levels, several limitations warrant consideration and provide directions for future research. The selection of the respective research methods naturally entails certain advantages and disadvantages.

The chosen methodological approaches (online and laboratory experiments) offer distinct advantages but also limitations. The online experiments conducted in *Paper 1* and *Paper 2* allowed for rapid data collection and access to larger, more diverse participant pools. However, these settings inherently lack control over environmental variables, introducing the risk of unobserved confounds that may have influenced participant behavior. Conversely, the laboratory studies in *Paper 3* and *Paper 4* enabled a high degree of experimental control, facilitating more reliable focus on the phenomenon under investigation by minimizing external noise. Yet, this simplification may limit external validity, as complex situational dynamics, further motivating factors, and additional hierarchies and responsibilities in organizational contexts are neglected.

Therefore, during the dissertation process, efforts were made to enhance the internal and external validity of the results. First, all experiments employed monetary incentives to ensure internal validity and to encourage participants to engage seriously with the tasks. This is a common approach to simulate real-world decision-making contexts and promote authentic behavior during the experimental sessions. Second, while this dissertation primarily emphasizes observable user behavior, such as reliance, error detection, or team dynamics, multiple attempts were made to incorporate internal psychological factors that may underlie these behaviors. For instance, participants' attitudes toward AI (Schepman and Rodway, 2020), perceived psychological safety (Edmondson, 1999), and individual personality traits (Rammstedt et al., 2014) were systematically assessed using validated scales. Nonetheless, it must be acknowledged that internal states are inherently difficult to access and interpret fully, and the omission of other potentially influential cognitive or affective factors limit the generalizability of the findings. Future research could, for example, attempt to quantify brain activity and flows or apply additional qualitative methods (e.g., the think-aloud method) to improve the operationalization of actual thought patterns and cognitive effort. Third, by focusing intentionally on non-expert users and student samples, this dissertation aims to reflect the growing relevance of these groups in everyday AI applications, especially as future decision-makers, given that non-expert users are increasingly collaborating with AI (Miller, 2019; Schoeffer et al., 2024; Liao et al., 2020). Moreover, students represent future professionals and managers who are likely to encounter AI-based systems in their future work environments (Zercher, Jussupow, and Heinzl, 2025). Nevertheless, the extent to which these samples can be generalized to experienced professionals in domain-specific or high-consequence environments is limited. Additionally, the studies in this dissertation are based on convenience-based sampling, where participants may have self-selected into the experiments. However, the control variables often indicate that the samples in this thesis behave similarly to other studies, suggesting at least some level of representativeness. However, more research is urgently needed to validate these results in field studies and real-world applications. Fourth, while attention was paid to achieving sample heterogeneity across studies (e.g., by recruiting participants across multiple institutions and countries in *Paper 4*), this approach cannot entirely account for potential cultural or contextual disparities. For instance, other studies indicate cultural differences in attitudes towards algorithms (Lee and Rich, 2021; Yam et al., 2023). Consequently, more research should investigate cross-cultural differences in the context of human-AI interaction.

A further key methodological consideration involves the limited observation periods of the studies included in this dissertation. All experiments were designed as short-term interactions, focusing on immediate behavioral responses to AI systems. While this approach allows this dissertation to investigate multiple variables under certain conditions, it does not account for the longitudinal dynamics of human-AI interaction. In real-world settings, users often engage with AI systems repeatedly, and such sustained exposure may lead to learning effects and different patterns of adaptation, trust calibration, or reliance behavior. Particularly in *Paper 4*, the permanent and temporary integration of AI into team dynamics implies distinct impacts on team processes and structure. This suggests that short-duration experiments may underestimate the systemic implications of AI collaboration over time.

Moreover, given the rapid pace of innovation and the sociotechnical uncertainty surrounding the future development of AI, it remains challenging to predict how human-AI interaction will evolve in the future. While previous research forecasts an era of automation (Frey and Osborne, 2017; Acemoglu and Restrepo, 2018), current literature hypothesizes a time of augmentation between humans and AI (Raisch and Krakowski, 2021; Raisch and Fomina, 2025; Brynjolfsson et al., 2025). Taking the different implications regarding the influence of familiarity with AI as an example, whereby on the one hand, initial skepticism toward AI can decrease through repeated use (Prahl and Van Swol, 2017; Jussupow et al., 2020); on the other hand, routine use and rapid advancements in AI can also foster uncritical acceptance of AI recommendations (Larson et al., 2024; Spitzer et al., 2024; Rebholz et al., 2024). This highlights the challenge of predicting developments in this highly disruptive and rapidly evolving domain. Therefore, longitudinal studies that track behavior across extended time frames, varied tasks, and real-world settings are crucial to capture the dynamic and adaptive nature of this relationship.

Ultimately, despite their temporal constraints, the studies presented in this dissertation offer a meaningful and timely window into how humans interact with AI systems under diverse conditions of task complexity, explanation design, decision context, and collaboration modality. By focusing on individual- and team-level interactions across multiple experimental paradigms, this dissertation provides a robust empirical foundation for future investigations.

1.5. Conclusion

In summary, this dissertation demonstrates that finding the balance between human intelligence and artificial intelligence, where they complement each other perfectly to achieve optimal synergy, is a genuinely challenging task. By examining this teamwork from different angles, applying different theoretical concepts, and evaluating heterogeneous influencing factors, this dissertation contributes to shedding light on this transformative academic field. This research highlights the complexity of developing AI systems that must not only be technically efficient but also cognitively and socially compatible with human users. Therefore, this thesis offers critical insights and approaches for further research on potential optimization strategies for human-AI interaction, including the influence of explanation modalities, the effects of accessibility to explanations, and response times. Although this research is often based on basic-level decision support or AI systems, the exploration schemes presented here can surely be adapted to the interaction with more complex AI systems. However, anticipating how different individuals will interact, behave, rely on, and collaborate with AI remains difficult. Human behavior is deeply nuanced and influenced by various factors, including personal factors such as expertise and thinking patterns, as well as contextual factors such as situational circumstances. These dynamics become even more unpredictable when layered with the rapid evolution of AI capabilities, leaving researchers to study a moving and co-constructed interaction space between humans and machines. As AI systems become more like teammates than tools, it is essential to move beyond static performance metrics and develop models that reflect the relational, dynamic, and co-adaptive qualities of intelligent collaboration. As this research illustrates, GenAI appears to alter not only what is discussed but also how communication is distributed among team members, prompting a reevaluation of power dynamics and leadership in AI-mediated teams. This dissertation reinforces the importance of ongoing, adaptive research that captures the fluid, co-evolving relationship between humans and AI.

References

Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., and Kankanhalli, M. (2018). Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, 1–18. https://doi.org/10.1145/3173574.3174156.

Acemoglu, D. and Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American economic review* (108:6), 1488–1542.

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

Agarwal, P. and Farndale, E. (2017). High-performance work systems and creativity implementation: the role of psychological capital and psychological safety. *Human Resource Management Journal* (27:3), 440–458.

Ågerfalk, P. J. (2020). Artificial intelligence as digital agency. *European Journal of Information Systems* (29:1), 1–8. https://doi.org/10.1080/0960085X.2020.1721947.

Agrawal, A., Gans, J., and Goldfarb, A. (2017). What to expect from artificial intelligence. *MIT Sloan Management Review Cambridge, MA* (58:3), 22–27.

Ajzen, I. (2001). Nature and operation of attitudes. *Annual review of psychology* (52:1), 27–58. https://doi.org/10.1146/annurev.psych.52.1.27.

Allcott, H. and Kessler, J. B. (2019). The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics* (11:1), 236–276.

Allcott, H. and Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review* (104:10), 3003–3037.

Amabile, T. M. (2018). Creativity in context: Update to the social psychology of creativity. *Routledge* (1). https://doi.org/10.4324/9780429501234.

Amabile, T. M. (2020). Creativity, artificial intelligence, and a world of surprises. *Academy of Management Discoveries* (6:3), 351–354.

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., et al. (2019). Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13.

Anantrasirichai, N. and Bull, D. (2022). Artificial intelligence in the creative industries: a review. *Artificial intelligence review* (55:1), 589–656. https://doi.org/10.1007/s10462-021-10039-7.

Araujo, T., Helberger, N., Kruikemeier, S., and de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY* (35:3), 611–623. https://doi.org/10.1007/s00146-019-00931-w.

Argote, L., Lee, S., and Park, J. (2021). Organizational learning processes and outcomes: Major findings and future research directions. *Management Science* (67:9), 5399–5429.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion* (58), 82–115. https://doi.org/10.1016/j.inffus.2019.12.012.

Ashkinaze, J., Mendelsohn, J., Qiwei, L., Budak, C., and Gilbert, E. (2024). How AI Ideas Affect the Creativity, Diversity, and Evolution of Human Ideas: Evidence From a Large, Dynamic Experiment. *arXiv preprint arXiv:2401.13481*.

Bader, S. and Oevermann, J. (2017). Semantic Annotation of Heterogeneous Data Sources: Towards an Integrated Information Framework for Service Technicians. In *Proceedings of the 13th International Conference on Semantic Systems*, 73–80.

Badue, C., Guidolini, R., Carneiro, R. V., Azevedo, P., Cardoso, V. B., Forechi, A., Jesus, L., Berriel, R., Paixao, T. M., Mutz, F., et al. (2021). Self-driving cars: A survey. *Expert Systems with Applications* (165), 113816.

Bailey, P. E., Leon, T., Ebner, N. C., Moustafa, A. A., and Weidemann, G. (2022). A meta-analysis of the weight of advice in decision-making. *Current Psychology* (42), 24516–24541. https://doi.org/10.1007/s12144-022-03573-2.

Baird, A. and Maruping, L. M. (2021). The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts. *MIS Q.* (45). URL: https://api.semanticscholar.org/CorpusID:228102761.

Balkundi, P. and Harrison, D. A. (2006). Ties, leaders, and time in teams: Strong inference about network structure's effects on team viability and performance. *Academy of Management Journal* (49:1), 49–68.

Bansal, G., Nushi, B., Kamar, E., Lasecki, W. S., Weld, D. S., and Horvitz, E. (2019). Beyond accuracy: The role of mental models in human-AI team performance. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* (7:1), 2–11.

Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T., and Weld, D. (2021). Does the whole exceed its parts? The effect of ai explanations on complementary team performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–16.

Bauer, K., von Zahn, M., and Hinz, O. (2023). Expl (AI) ned: The impact of explainable artificial intelligence on users' information processing. *Information Systems Research* (34:4), 1582–1602. https://doi.org/10.1287/isre.2023.1199.

Bawden, D. and Robinson, L. (2009). The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of Information Science* (35:2), 180–191.

Bell, A. and Jones, K. (2015). Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods* (3:1), 133–153.

Beyth-Marom, R. (1982). How probable is probable? A numerical translation of verbal probability expressions. *Journal of Forecasting* (1:3), 257–269.

Bicchieri, C. and Dimant, E. (2022). Nudging with care: The risks and benefits of social information. *Public choice* (191:3), 443–464.

Bohren, N., Hakimov, R., and Lalive, R. (2024). *Creative and Strategic Capabilities of Generative AI: Evidence from Large-Scale Experiments*. Tech. rep. Institut of Labor Economics (IZA). URL: https://hdl.handle.net/10419/305744.

Bonaccio, S. and Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes* (101:2), 127–151.

Bonnefon, J.-F. and Rahwan, I. (2020). Machine thinking, fast and slow. *Trends in Cognitive Sciences* (24:12), 1019–1027.

Booch, G., Fabiano, F., Horesh, L., Kate, K., Lenchner, J., Linck, N., Loreggia, A., Murugesan, K., Mattei, N., Rossi, F., and Srivastava, B. (2021). Thinking Fast and Slow in AI. In *Proceedings of the AAAI Conference on Artificial Intelligence* (35:17), 15042–15046. https://doi.org/10.1609/aaai.v35i17.17765.

Bouschery, S. G., Blazevic, V., and Piller, F. T. (2023). Augmenting human innovation teams with artificial intelligence: Exploring transformer-based language models. *Journal of Product Innovation Management* (40:2), 139–153. https://doi.org/10.1111/jpim.12656.

Boussioux, L., Lane, J. N., Zhang, M., Jacimovic, V., and Lakhani, K. R. (2024). The Crowdless Future? Generative AI and Creative Problem-Solving. *Organization Science* (35:5), 1589–1607. https://doi.org/10.1287/orsc.2023.18430.

Bradler, C., Neckermann, S., and Warnke, A. J. (2016). Incentivizing creativity: A large-scale experiment with tournaments and gifts. *ZEW-Centre for European Economic Research Discussion Paper*.

Brüderl, J. and Ludwig, V. (2015). Fixed-effects panel regression. *The Sage handbook of regression analysis and causal inference* (327), 357.

Brynjolfsson, E., Li, D., and Raymond, L. (2025). Generative AI at work. *The Quarterly Journal of Economics* (140:2), 889–942.

Brynjolfsson, E., Mitchell, T., and Rock, D. (2018). What can machines learn and what does it mean for occupations and the economy? In *AEA papers and proceedings* (108), 43–47.

Buçinca, Z., Malaya, M. B., and Gajos, K. Z. (2021). To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-computer Interaction* (5:CSCW1), 1–21.

Budescu, D. V. and Wallsten, T. S. (1985). Consistency in interpretation of probabilistic phrases. *Organizational Behavior and Human Decision Processes* (36:3), 391–405.

Bunt, A., Lount, M., and Lauzon, C. (2012). Are explanations always important? A study of deployed, low-cost intelligent interactive systems. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, 169–178.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society* (3:1). https://doi.org/10.1177/2053951715622512.

Burton, J. W., Stein, M.-K., and Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making* (33:2), 220–239.

Byrne, R. M. (2019). Counterfactuals in Explainable Artificial Intelligence (XAI): Evidence from Human Reasoning. In *IJCAI*, 6276–6282.

Caluori, L. (2024). Hey Alexa, why are you called intelligent? An empirical investigation on definitions of AI. *AI & SOCIETY* (39:4), 1905–1919.

Carton, S., Mei, Q., and Resnick, P. (2020). Feature-based explanations don't help people detect misclassifications of online toxicity. In *Proceedings of the International AAAI Conference on Web and Social Media* (14), 95–106.

Castelo, N., Bos, M. W., and Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research* (56:5), 809–825.

Chen, C. X., Williamson, M. G., and Zhou, F. H. (2012). Reward system design and group creativity: An experimental investigation. *The Accounting Review* (87:6), 1885–1911.

Chen, D. L., Schonger, M., and Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance* (9), 88–97.

Chen, Z. and Chan, J. (2024). Large language model in creative work: The role of collaboration modality and user expertise. *Management Science* (70:12), 9101–9117.

Cheng, H.-F., Wang, R., Zhang, Z., O'Connell, F., Gray, T., Harper, F. M., and Zhu, H. (2019). Explaining decision-making algorithms through UI: Strategies to help non-expert stakeholders. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12.

Chiang, C.-W. and Yin, M. (2022). Exploring the Effects of Machine Learning Literacy Interventions on Laypeople's Reliance on Machine Learning Models. In *27th International Conference on Intelligent User Interfaces*, 148–161.

Chu, E., Roy, D., and Andreas, J. (2020). Are Visual Explanations Useful? A Case Study in Model-in-the-Loop Prediction. *arXiv preprint arXiv: 2007.12248*.

Clark, E., Ross, A. S., Tan, C., Ji, Y., and Smith, N. A. (2018). Creative writing with a machine in the loop: Case studies on slogans and stories. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces*, 329–340.

Contractor, N. S., Wasserman, S., and Faust, K. (2006). Testing multitheoretical, multilevel hypotheses about organizational networks: An analytic framework and empirical example. *Academy of Management Review* (31:3), 681–703.

Craik, K. J. W. (1967). *The nature of explanation*. CUP Archive.

Cummings, J. N. and Cross, R. (2003). Structural properties of work groups and their consequences for performance. *Social Networks* (25:3), 197–210.

David, L., Vassena, E., and Bijleveld, E. (2024). The unpleasantness of thinking: A meta-analytic review of the association between mental effort and negative affect. *Psychological Bulletin* (150:9), 1070–1093. https://doi.org/10.1037/bul0000443.

De Freitas, J., Oğuz-Uğuralp, Z., Uğuralp, A. K., and Puntoni, S. (2025). AI companions reduce loneliness. *Journal of Consumer Research*.

De Graaf, M. M. and Malle, B. F. (2017). How people explain action (and autonomous intelligent systems should too). In *AAAI Fall Symposium Series* (8).

De Winter, J. C. and Dodou, D. (2014). Why the Fitts list has persisted throughout the history of function allocation. *Cognition, Technology & Work* (16:1), 1–11.

Dell'Acqua, F., Ayoubi, C., Lifshitz, H., Sadun, R., Mollick, E., Mollick, L., Han, Y., Goldman, J., Nair, H., Taub, S., et al. (2025). *The cybernetic teammate: a field experiment on generative AI reshaping teamwork and expertise*. Tech. rep. National Bureau of Economic Research.

Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., and Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper 24–013*.

Demir, S., Fuegener, A., Gupta, A., and Weinmann, M. (2024). When AI is Creative: How Do Humans Perceive Creativity When AI is Involved? In *ICIS 2024 Proceedings*.

Diana, M. and Marescaux, J. (2015). Robotic surgery. *British Journal of Surgery* (102:2), e15–e28.

Diehl, M. and Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology* (53:3), 497.

Dietvorst, B. J., Massey, C., and Simmons, J. P. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* (144:1), 114.

Dietvorst, B. J., Simmons, J. P., and Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science* (64:3), 1155–1170.

Doshi, A. R. and Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances* (10:28), eadn5290.

Doshi-Velez, F. and Kim, B. (2017). Towards a Rigorous Science of Interpretable Machine Learning. *arXiv preprint arXiv:1702.08608*.

Douglas, L., Jillian, M., Thomas, L., and Eric, L. (2006). Identifying quality, novel, and creative ideas: constructs and scales for idea evaluation. *Journal of the Association for Information Systems* (7:10), 646.

Dugosh, K. L., Paulus, P. B., Roland, E. J., and Yang, H.-C. (2000). Cognitive stimulation in brainstorming. *Journal of Personality and Social Psychology* (79:5), 722.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., and Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies* (58:6), 697–718.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly* (44:2), 350–383. https://doi.org/10.2307/2666999.

Edmondson, A. C. and Bransby, D. P. (2023). Psychological safety comes of age: Observed themes in an established literature. *Annual Review of Organizational Psychology and Organizational Behavior* (10:1), 55–78.

Edmondson, A. C. and Lei, Z. (2014). Psychological safety: The history, renaissance, and future of an interpersonal construct. *Annual Review of Organizational Psychology and Organizational Behavior* (1:1), 23–43.

Efendić, E., Van de Calseyde, P. P., and Evans, A. M. (2020). Slow response times undermine trust in algorithmic (but not human) predictions. *Organizational Behavior and Human Decision Processes* (157), 103–114.

Ehrlich, K., Kirk, S. E., Patterson, J., Rasmussen, J. C., Ross, S. I., and Gruen, D. M. (2011). Taking advice from intelligent systems: the double-edged sword of explanations. In *Proceedings of the 16th International Conference on Intelligent User Interfaces*, 125–134.

Enholm, I. M., Papagiannidis, E., Mikalef, P., and Krogstie, J. (2022). Artificial intelligence and business value: A literature review. *Information Systems Frontiers* (24:5), 1709–1734.

Epstein, Z., Hertzmann, A., Herman, L. M., Mahari, R., Frank, M. R., Groh, M., Schroeder, H., Smith, A., Akten, M., Fjeld, J., Farid, H., Leach, N., Pentland, A., and Russakovsky, O. (2023). Art and the science of generative AI. *Science* (380), 1110–1111. https://doi.org/10.1126/science.adh4451.

Erev, I. and Cohen, B. L. (1990). Verbal versus numerical probabilities: Efficiency, biases, and the preference paradox. *Organizational Behavior and Human Decision Processes* (45:1), 1–18. https://doi.org/10.1016/0749-5978(90)90002-Q.

EU (2021). Proposal for a regulation EU of the European Parliament and of the Council of April 21, laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. *Official J. Eur. Union Law* (119).

Fast, E. and Horvitz, E. (2017). Long-term trends in the public perception of artificial intelligence. In *Proceedings of the AAAI Conference on Artificial Intelligence* (31:1). https://doi.org/10.1609/aaai.v31i1.10635.

Fitzsimons, G. J. and Lehmann, D. R. (2004). Reactance to recommendations: When unsolicited advice yields contrary responses. *Marketing Science* (23:1), 82–94.

Frey, C. B. and Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change* (114), 254–280.

Fu, R., Aseri, M., Singh, P. V., and Srinivasan, K. (2022). "Un" fair machine learning algorithms. *Management Science* (68:6), 4173–4195. https://doi.org/10.1287/mnsc.2021.4065.

Fügener, A., Grahl, J., Gupta, A., and Ketter, W. (2021). Will humans-in-the-loop become borgs? Merits and pitfalls of working with AI. *Management Information Systems Quarterly* (45).

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

Gajos, K. Z. and Mamykina, L. (2022). Do people engage cognitively with AI? Impact of AI assistance on incidental learning. In *27th International Conference on Intelligent User Interfaces*, 794–806. https://doi.org/10.1145/3490099.3511138.

Gaudiello, I., Zibetti, E., Lefort, S., Chetouani, M., and Ivaldi, S. (2016). Trust as indicator of robot functional and social acceptance. An experimental study on user conformation to iCub answers. *Computers in Human Behavior* (61), 633–655.

Gerlings, J., Jensen, M. S., and Shollo, A. (2021). Explainable AI, but explainable to whom? An exploratory case study of xAI in healthcare. In *Handbook of Artificial Intelligence in Healthcare: Vol 2: Practicalities and Prospects*. Springer, 169–198.

Gibson, D. R. (2005). Taking Turns and Talking Ties: Networks and Conversational Interaction. *American Journal of Sociology* (110:6), 1561–1597. https://doi.org/10.1086/428689.

Gigerenzer, G., Hertwig, R., Van Den Broek, E., Fasolo, B., and Katsikopoulos, K. V. (2005). "A 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts? *Risk Analysis: An International Journal* (25:3), 623–629.

Gil, D., Hobson, S., Mojsilović, A., Puri, R., and Smith, J. R. (2020). AI for Management: An Overview. In *The Future of Management in an AI World*. Springer, 3–19.

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., and Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, 80–89.

Gino, F., Brooks, A. W., and Schweitzer, M. E. (2012). Anxiety, advice, and the ability to discern: feeling anxious motivates individuals to seek and use advice. *Journal of Personality and Social Psychology* (102:3), 497. https://doi.org/10.1037/a0026413.

Gino, F. and Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making* (20:1), 21–35. https://doi.org/10.1002/bdm.539.

Giray, L. (2024). The Problem with False Positives: AI Detection Unfairly Accuses Scholars of AI Plagiarism. *The Serials Librarian* (85:5-6), 181–189. https://doi.org/10.1080/0361526X.2024.2433256.

Glikson, E. and Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals* (14:2), 627–660.

Gneiting, T. and Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application* (1), 125–151.

Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press.

Goguen, J. A., Weiner, J., and Linde, C. (1983). Reasoning and natural explanation. *International Journal of Man-Machine Studies* (19:6), 521–559.

Greenbaum, R. L., Bonner, J. M., Mawritz, M. B., Butts, M. M., and Smith, M. B. (2020). It is all about the bottom line: Group bottom-line mentality, psychological safety, and group creativity. *Journal of Organizational Behavior* (41:6), 503–517.

Gregor, S. and 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.

Grilli, L. and Pedota, M. (2024). Creativity and artificial intelligence: A multilevel perspective. *Creativity and Innovation Management* (33:2), 234–247.

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

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys (CSUR)* (51:5), 1–42. https://doi.org/10.1145/3236009.

Gunning, D. (2017). Explainable artificial intelligence (xai). *Defense advanced research projects agency (DARPA)* (2:2), 1.

Gunning, D. and Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI magazine* (40:2), 44–58.

Guzman, A. L. (2020). Ontological boundaries between humans and computers and the implications for human-machine communication. *Human-Machine Communication* (1), 37–54.

Guzman, A. L. and Lewis, S. C. (2020). Artificial intelligence and communication: A human–machine communication research agenda. *New Media & Society* (22:1), 70–86.

Haefner, N., Wincent, J., Parida, V., and Gassmann, O. (2021). Artificial intelligence and innovation management: A review, framework, and research agenda. *Technological Forecasting and Social Change* (162), 120392.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., and Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors* (53:5), 517–527. https://doi.org/10.1177/0018720811417254.

Hecker, S., Dai, D., and Van Gool, L. (2018). Failure prediction for autonomous driving. In *2018 IEEE Intelligent Vehicles Symposium (IV)*, 1792–1799.

Hennessey, B. A., Amabile, T. M., and Mueller, J. S. (2011). Consensual assessment. In *Encyclopedia of Creativity*. M. A. Runco and S. R. Pritzker (eds.). San Diego: Academic Press, 253–260.

Henttonen, K. (2010). Exploring social networks on the team level—A review of the empirical literature. *Journal of Engineering and Technology Management* (27:1-2), 74–109.

Hinton, G. (2024). *Banquet Speech*. NobelPrize.org. Nobel Prize Outreach 2025. URL: https://www.nobelprize.org/prizes/physics/2024/hinton/speech/.

Hoff, K. A. and Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors* (57:3), 407–434.

Hofheinz, C., Germar, M., Schultze, T., Michalak, J., and Mojzisch, A. (2017). Are Depressed People More or Less Susceptible to Informational Social Influence? *Cognitive Therapy and Research* (41), 699–711. https://doi.org/10.1007/s10608-017-9848-7.

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., and Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (9:4), e1312. https://doi.org/10.1002/widm.1312.

Hora, S., Lemoine, G. J., Xu, N., and Shalley, C. E. (2021). Unlocking and closing the gender gap in creative performance: A multilevel model. *Journal of Organizational Behavior* (42:3), 297–312.

Horstmann, A. C., Bock, N., Linhuber, E., Szczuka, J. M., Straßmann, C., and Krämer, N. C. (2018). Do a robot's social skills and its objection discourage interactants from switching the robot off? *PloS one* (13:7), e0201581. https://doi.org/10.1371/journal.pone.0201581.

Hou, Y. T.-Y. and Jung, M. F. (2021). Who is the expert? Reconciling algorithm aversion and algorithm appreciation in AI-supported decision making. *Proceedings of the ACM on Human-Computer Interaction* (5:CSCW2), 1–25.

Hubner, S., Baum, M., and Frese, M. (2020). Contagion of entrepreneurial passion: Effects on employee outcomes. *Entrepreneurship Theory and Practice* (44:6), 1112–1140.

Hull, C. L. (1943). *Principles of behavior: An introduction to behavior therapy*. Appleton Century Crofts.

Johnson, E. J. and Payne, J. W. (1985). Effort and accuracy in choice. *Management Science* (31:4), 395–414. https://doi.org/10.1287/mnsc.31.4.395.

Johnson, P. E. (1983). What kind of expert should a system be? *The Journal of Medicine and Philosophy* (8:1), 77–97. https://doi.org/10.1093/jmp/8.1.77.

Jussupow, E., Benbasat, I., and Heinzl, A. (2020). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. In *Proceedings of the 28th European Conference on Information Systems (ECIS)*, 1–16.

Jussupow, E., Spohrer, K., Heinzl, A., and Gawlitza, J. (2021). Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence. *Information Systems Research* (32:3), 713–735.

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

Kahneman, D., Sibony, O., and Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Hachette UK.

Kallinikos, J. and Hasselbladh, H. (2009). Work, control and computation: Rethinking the legacy of neo-institutionalism. In *Institutions and Ideology, Research in the Sociology of Organizations*. R. E. Meyer, K. Sahlin, M. J. Ventresca, and P. Walgenbach (eds.). 27. Emerald Group Publishing Limited, 257–282. https://doi.org/10.1108/S0733-558X(2009)000002701.

Kauffeld, S. and Lehmann-Willenbrock, N. (2012). Meetings Matter: Effects of Team Meetings on Team and Organizational Success. *Small Group Research* (43:2), 130–158. https://doi.org/10.1177/1046496411429599.

Kauffeld, S., Lehmann-Willenbrock, N., and Meinecke, A. L. (2018). The Advanced Interaction Analysis for Teams (act4teams) Coding Scheme. In. *The Cambridge Handbook of Group Interaction Analysis*. Cambridge Handbooks in Psychology. Cambridge University Press, 422–431.

Kim, J.-Y. and Heo, W. (2021). Artificial intelligence video interviewing for employment: perspectives from applicants, companies, developer and academicians. *Information Technology & People* (35:3), 861–878.

Köbis, N. and Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior* (114), 106553.

Kohn, N. W. and Smith, S. M. (2011). Collaborative fixation: Effects of others' ideas on brainstorming. *Applied Cognitive Psychology* (25:3), 359–371.

Kohn, S. C., De Visser, E. J., Wiese, E., Lee, Y.-C., and Shaw, T. H. (2021). Measurement of trust in automation: A narrative review and reference guide. *Frontiers in Psychology* (12), 604977.

Kouagou, N., Heindorf, S., Demir, C., and Ngomo, A.-C. N. (2022). Learning Concept Lengths Accelerates Concept Learning in ALC. In *European Semantic Web Conference*, 236–252.

Koudenburg, N., Postmes, T., and Gordijn, E. H. (2017). Beyond content of conversation: The role of conversational form in the emergence and regulation of social structure. *Personality and Social Psychology Review* (21:1), 50–71.

Kovács, B. (2024). The Turing test of online reviews: Can we tell the difference between human-written and GPT-4-written online reviews? *Marketing Letters* (35:4), 651–666.

Kozlowski, S. W. and Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest* (7:3), 77–124.

Kunda, Z. (1999). *Social cognition: Making sense of people*. MIT press.

Labianca, G., Brass, D. J., and Gray, B. (1998). Social networks and perceptions of intergroup conflict: The role of negative relationships and third parties. *Academy of Management Journal* (41:1), 55–67.

Lai, V. and Tan, C. (2019). On human predictions with explanations and predictions of machine learning models: A case study on deception detection. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 29–38.

Larson, B. Z., Moser, C., Caza, A., Muehlfeld, K., and Colombo, L. A. (2024). Critical Thinking in the Age of Generative AI. *Academy of Management Learning & Education* (23:3), 373–378. https://doi.org/10.5465/amle.2024.0338.

Laske, K. and Schroeder, M. (2017). Quantity, quality and originality: The effects of incentives on creativity. In *Beiträge zur Jahrestagung des Vereins fürr Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Organization and Firms II*.

Laupichler, M. C., Aster, A., Schirch, J., and Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence* (3), 100101. https://doi.org/10.1016/j.caeai.2022.100101.

Lebedeva, A., Kornowicz, J., Lammert, O., and Papenkordt, J. (2023). The Role of Response Time for Algorithm Aversion in Fast and Slow Thinking Tasks. In *International Conference on Human-Computer Interaction*, 131–149. https://doi.org/10.1007/978-3-031-35891-3_9.

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., and Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–22.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society* (5:1), 2053951718756684.

Lee, M. K. and Rich, K. (2021). Who is included in human perceptions of AI?: Trust and perceived fairness around healthcare AI and cultural mistrust. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3411764.3445570.

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., and Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology* (31:4), 611–627.

Lewicki, R. J. (1996). Developing and maintaining trust in work relationships. *Trust in organizations: Frontiers of theory and research/Sage*, 114–139.

Li, H., Zhang, R., Lee, Y.-C., Kraut, R. E., and Mohr, D. C. (2023). Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. *NPJ Digital Medicine* (6:1), 236.

Liao, Q. V., Gruen, D., and Miller, S. (2020). Questioning the AI: informing design practices for explainable AI user experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15.

Liel, Y. and Zalmanson, L. (2023). Turning Off Your Better Judgment – Conformity to Algorithmic Recommendations. *Academy of Management Proceedings* (2023:1), 16319. https://doi.org/10.5465/AMPROC.2023.277bp.

Lipkus, I. M. (2007). Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. *Medical Decision Making* (27:5), 696–713.

Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* (16:3), 31–57.

Liu, Y. and Keller, R. T. (2021). How psychological safety impacts R&D project teams' performance. *Research-Technology Management* (64:2), 39–45.

Loftus, E. F., Miller, D. G., and Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory* (4:1), 19.

Logg, J. M., Minson, J. A., and Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes* (151), 90–103.

Long, D. and Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16.

Lu, T. and Zhang, Y. (2024). 1+ 1> 2? Information, humans, and machines. *Information Systems Research* (36:1), 394–418.

Mahmud, H., Islam, A. N., Ahmed, S. I., and Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change* (175), 121390.

Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures* (90), 46–60.

Marks, M. A., Mathieu, J. E., and Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review* (26:3), 356–376.

Martens, D., Baesens, B., Van Gestel, T., and Vanthienen, J. (2007). Comprehensible credit scoring models using rule extraction from support vector machines. *European Journal of Operational Research* (183:3), 1466–1476.

Mathieu, J. E., Gallagher, P. T., Domingo, M. A., and Klock, E. A. (2019). Embracing complexity: Reviewing the past decade of team effectiveness research. *Annual Review of Organizational Psychology and Organizational Behavior* (6:1), 17–46.

Mathieu, J. E., Luciano, M. M., D'Innocenzo, L., Klock, E. A., and LePine, J. A. (2020). The development and construct validity of a team processes survey measure. *Organizational Research Methods* (23:3), 399–431.

Mauboussin, A. and Mauboussin, M. J. (2018). If you say something is "likely," how likely do people think it is. *Harvard Business Review* (3:7).

McBride, M., Carter, L., and Ntuen, C. (2012). The impact of personality on nurses' bias towards automated decision aid acceptance. *International Journal of Information Systems and Change Management* (6:2), 132–146.

McGrath, J. E. (1964). Social psychology: A brief introduction. *NY: Holt, Rinehart and Winston*.

McNeese, N. J., Demir, M., Cooke, N. J., and Myers, C. (2018). Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human Factors* (60:2), 262–273.

Mikalef, P., Conboy, K., Lundström, J. E., and Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side'of AI. *European Journal of Information Systems* (31:3), 257–268.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence* (267), 1–38.

Miller, T. (2023). Explainable ai is dead, long live explainable ai! hypothesis-driven decision support using evaluative ai. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 333–342.

Millet, K., Buehler, F., Du, G., and Kokkoris, M. D. (2023). Defending humankind: Anthropocentric bias in the appreciation of AI art. *Computers in Human Behavior* (143), 107707. https://doi.org/10.1016/j.chb.2023.107707.

Mislavsky, R. and Gaertig, C. (2022). Combining probability forecasts: 60% and 60% is 60%, but likely and likely is very likely. *Management Science* (68:1), 541–563.

Mittelstadt, B., Russell, C., and Wachter, S. (2019). Explaining explanations in AI. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 279–288.

Mohseni, S., Zarei, N., and Ragan, E. D. (2021). A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* (11:3-4), 1–45.

Moon, Y. (2000). Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of Consumer Research* (26:4), 323–339.

Moorthy, K., Munz, Y., Dosis, A., Hernandez, J., Martin, S., Bello, F., Rockall, T., and Darzi, A. (2004). Dexterity enhancement with robotic surgery. *Surgical Endoscopy and Other Interventional Techniques* (18), 790–795.

Murray, A., Rhymer, J., and Sirmon, D. G. (2021). Humans and technology: Forms of conjoined agency in organizations. *Academy of Management Review* (46:3), 552–571.

Neef, N. E., Zabel, S., Papoli, M., and Otto, S. (2024). Drawing the full picture on diverging findings: Adjusting the view on the perception of art created by Artificial Intelligence. *AI & SOCIETY*. https://doi.org/10.1007/s00146-024-02020-z.

Ng, D. T. K., Leung, J. K. L., Chu, K. W. S., and Qiao, M. S. (2021a). AI literacy: Definition, teaching, evaluation and ethical issues. *Proceedings of the Association for Information Science and Technology* (58:1), 504–509.

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., and Qiao, M. S. (2021b). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence* (2), 100041.

Nijstad, B. A. and Stroebe, W. (2006). How the group affects the mind: A cognitive model of idea generation in groups. *Personality and Social Psychology Review* (10:3), 186–213.

Nijstad, B. A., Stroebe, W., and Lodewijkx, H. F. (2003). Production blocking and idea generation: Does blocking interfere with cognitive processes? *Journal of Experimental Social Psychology* (39:6), 531–548.

Noy, S. and Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science* (381:6654), 187–192.

O'Neill, T. A., Flathmann, C., McNeese, N. J., and Salas, E. (2023). Human-autonomy Teaming: Need for a guiding team-based framework? *Computers in Human Behavior* (146), 107762.

O'Neill, T., McNeese, N., Barron, A., and Schelble, B. (2022). Human–autonomy teaming: A review and analysis of the empirical literature. *Human Factors* (64:5), 904–938.

OpenAI (2023). *GPT-4 Technical Report*. Technical Report. Accessed: 2025-07-22. OpenAI. URL: https://cdn.openai.com/papers/gpt-4.pdf.

Papenkordt, J. (2024). Navigating Transparency: The Influence of On-demand Explanations on Non-expert User Interaction with AI. In *International Conference on Human-Computer Interaction*, 238–263. https://doi.org/10.1007/978-3-031-60606-9_14.

Papenkordt, J., Dahlke, J., Neef, N., and Zabel, S. (2025). *Exploring the impact of AI on team collaboration dynamics in creative decision-making*. Working Papers Dissertations 146. Paderborn University, Faculty of Business Administration and Economics, 2025-07. URL: https://ideas.repec.org/p/pdn/dispap/146.html.

Papenkordt, J., Ngonga Ngomo, A.-C., and Thommes, K. (2023). Are Numbers or Words the Key to User Reliance on AI? In *Academy of Management Proceedings* (2023:1), 12946.

Papenkordt, J., Ngonga Ngomo, A.-C., and Thommes, K. (2025). *Are numerical or verbal explanations of AI the key to appropriate user reliance and error detection? An experimental study* *with a classification algorithm*. Working Papers Dissertations 147. Paderborn University, Faculty of Business Administration and Economics, 2025-07. URL: https://ideas.repec.org/p/pdn/dispap/147.html.

Park, J. S., Barber, R., Kirlik, A., and Karahalios, K. (2019). A slow algorithm improves users' assessments of the algorithm's accuracy. *Proceedings of the ACM on Human-Computer Interaction* (3:CSCW), 1–15.

Paulus, P. (2000). Groups, Teams, and Creativity: The Creative Potential of Idea-generating Groups. *Applied Psychology* (49:2), 237–262. https://doi.org/10.1111/1464-0597.00013.

Paulus, P. B. and Nijstad, B. A. (2019). *The Oxford handbook of group creativity and innovation*. Oxford University Press.

Pedreschi, D., Giannotti, F., Guidotti, R., Monreale, A., Ruggieri, S., and Turini, F. (2019). Meaningful explanations of black box AI decision systems. In *Proceedings of the AAAI Conference on Artificial Intelligence* (33), 9780–9784.

Peng, S., Kalliamvakou, E., Cihon, P., and Demirer, M. (2023). *The impact of ai on developer productivity: Evidence from github copilot*. arXiv: 2302.06590 [cs.SE].

Pinsonneault, A., Barki, H., Gallupe, R. B., and Hoppen, N. (1999). Electronic brainstorming: The illusion of productivity. *Information Systems Research* (10:2), 110–133.

Plumb, G., Molitor, D., and Talwalkar, A. S. (2018). Model agnostic supervised local explanations. *Advances in neural information processing systems* (31).

Posit team (2025). *RStudio: Integrated Development Environment for R*. Posit Software, PBC. Boston, MA. URL: http://www.posit.co/.

Poursabzi-Sangdeh, F., Goldstein, D. G., Hofman, J. M., Wortman Vaughan, J. W., and Wallach, H. (2021). Manipulating and measuring model interpretability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–52.

Poushneh, A. (2021). Humanizing voice assistant: The impact of voice assistant personality on consumers' attitudes and behaviors. *Journal of Retailing and Consumer Services* (58), 102283.

Prahl, A. and Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted? *Journal of Forecasting* (36:6), 691–702.

Proctor, T. (2020). Creative problem-solving techniques, paradigm shift and team performance. *Team Performance Management: An International Journal* (26:7/8), 451–466.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., et al. (2019). Machine behaviour. *Nature* (568:7753), 477–486.

Raisch, S. and Fomina, K. (2025). Combining human and artificial intelligence: Hybrid problem-solving in organizations. *Academy of Management Review* (50:2), 441–464.

Raisch, S. and Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review* (46:1), 192–210.

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., and Kovaleva, A. (2014). Big Five Inventory (BFI-10). de. *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. https://doi.org/10.6102/ZIS76. URL: https://zis.gesis.org/DoiId/zis76.

Raymond, M. and Rousset, F. (1995). An exact test for population differentiation. *Evolution* (49:6), 1280–1283.

Rebholz, T. R., Koop, A., and Hütter, M. (2024). Conversational user interfaces: Explanations and interactivity positively influence advice taking from generative artificial intelligence. *Technology, Mind, and Behavior* (5:4).

Ribeiro, M. T., Singh, S., and Guestrin, C. (2018). Anchors: High-precision model-agnostic explanations. In *Proceedings of the AAAI Conference on Artificial Intelligence* (32:1).

Rip, A. (2019). Folk theories of nanotechnologists. In *Nanotechnology and Its Governance*. Routledge, 56–74.

Rosenfeld, A. and Richardson, A. (2019). Explainability in human–agent systems. *Autonomous Agents and Multi-Agent Systems* (33), 673–705.

Rossi, F. and Loreggia, A. (2019). Preferences and Ethical Priorities: Thinking Fast and Slow in AI. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, 3–4.

Salas, E., Reyes, D. L., and McDaniel, S. H. (2018). The science of teamwork: Progress, reflections, and the road ahead. *American Psychologist* (73:4), 593.

Sauer, N. C. and Kauffeld, S. (2016). The structure of interaction at meetings: A social network analysis. *Zeitschrift für Arbeits-und Organisationspsychologie A&O*.

Scharowski, N., Perrig, S. A., von Felten, N., and Brühlmann, F. (2022). Trust and Reliance in XAI–Distinguishing Between Attitudinal and Behavioral Measures. *arXiv preprint arXiv:2203.12318*.

Schemmer, M., Hemmer, P., Nitsche, M., Kühl, N., and Vössing, M. (2022). A meta-analysis of the utility of explainable artificial intelligence in human-AI decision-making. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 617–626. https://doi.org/10.1145/3514094.3534128.

Schepman, A. and Rodway, P. (2020). Initial validation of the general attitudes towards Artificial Intelligence Scale. *Computers in Human Behavior Reports* (1), 100014.

Schepman, A. and Rodway, P. (2023). The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory validation and associations with personality, corporate distrust, and general trust. *International Journal of Human–Computer Interaction* (39:13), 2724–2741.

Schmidt, P., Biessmann, F., and Teubner, T. (2020). Transparency and trust in artificial intelligence systems. *Journal of Decision Systems* (29:4), 260–278.

Schoeffer, J., De-Arteaga, M., and Kuehl, N. (2024). Explanations, fairness, and appropriate reliance in human-AI decision-making. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–18.

Schoonderwoerd, T. A., Jorritsma, W., Neerincx, M. A., and Van Den Bosch, K. (2021). Human-centered XAI: Developing design patterns for explanations of clinical decision support systems. *International Journal of Human-Computer Studies* (154), 102684.

Schuetz, S. and Venkatesh, V. (2020). The rise of human machines: How cognitive computing systems challenge assumptions of user-system interaction. *Journal of the Association for Information Systems* (21:2), 460–482.

Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., and Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information Management* (57:2), 103174. https://doi.org/10.1016/j.im.2019.103174.

Sharan, N. N. and Romano, D. M. (2020). The effects of personality and locus of control on trust in humans versus artificial intelligence. *Heliyon* (6:8), e04572. https://doi.org/10.1016/j.heliyon.2020.e04572.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., and Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience* (40:1), 99–124.

Shrout, P. E. and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* (86:2), 420.

Simon, S., Sankaranarayanan, S., Tajik, E., Borchers, C., Shahrokhian, B., Balzan, F., Strauß, S., Viswanathan, S. A., Ataş, A. H., Čarapina, M., Liang, L., and Celik, B. (2025). Comparing a Human's and a Multi-Agent System's Thematic Analysis: Assessing Qualitative Coding Consistency. In *Artificial Intelligence in Education*, 60–73. https://doi.org/10.1007/978-3-031-98420-4_5.

Sjøvold, E., Olsen, T. R., and Heldal, F. (2022). Use of Technology in the Study of Team-Interaction and Performance. *Small Group Research* (53:4), 596–630. https://doi.org/10.1177/10464964211069328.

Skjuve, M., Følstad, A., Fostervold, K. I., and Brandtzaeg, P. B. (2022). A longitudinal study of human–chatbot relationships. *International Journal of Human-Computer Studies* (168), 102903.

Sparrowe, R. T., Liden, R. C., Wayne, S. J., and Kraimer, M. L. (2001). Social networks and the performance of individuals and groups. *Academy of Management Journal* (44:2), 316–325.

Spitzer, P., Holstein, J., Morrison, K., Holstein, K., Satzger, G., and Kühl, N. (2024). Don't be Fooled: The Misinformation Effect of Explanations in Human-AI Collaboration. *arXiv preprint arXiv:2409.12809*.

Strich, F., Mayer, A.-S., and Fiedler, M. (2021). What do I do in a world of artificial intelligence? Investigating the impact of substitutive decision-making AI systems on employees' professional role identity. *Journal of the Association for Information Systems* (22:2), 9.

Stuart, H. C. (2017). Structural disruption, relational experimentation, and performance in professional hockey teams: A network perspective on member change. *Organization Science* (28:2), 283–300.

Tambe, P., Cappelli, P., and Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review* (61:4), 15–42.

Thiebes, S., Lins, S., and Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets* (31:2), 447–464.

Thurman, N., Moeller, J., Helberger, N., and Trilling, D. (2019). My friends, editors, algorithms, and I: Examining audience attitudes to news selection. *Digital Journalism* (7:4), 447–469.

Turing, A. (2004). Intelligent machinery (1948). *B. Jack Copeland*, 395.

Vaccaro, M., Almaatouq, A., and Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour* (8:12), 2293–2303.

Van der Waa, J., Nieuwburg, E., Cremers, A., and Neerincx, M. (2021). Evaluating XAI: A comparison of rule-based and example-based explanations. *Artificial Intelligence* (291), 103404.

Varshney, K. R. and Alemzadeh, H. (2017). On the safety of machine learning: Cyber-physical systems, decision sciences, and data products. *Big Data* (5:3), 246–255.

Vasconcelos, H., Jörke, M., and Grunde-McLaughlin, M. (2022). When Do XAI Methods Work? A Cost-Benefit Approach to Human-AI Collaboration. In *CHI Workshop on Trust and Reliance in AI-Human Teams*. URL: https://api.semanticscholar.org/CorpusID:253387060.

Vilone, G. and Longo, L. (2021). Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion* (76), 89–106.

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

Walliser, J. C., de Visser, E. J., Wiese, E., and Shaw, T. H. (2019). Team structure and team building improve human–machine teaming with autonomous agents. *Journal of Cognitive Engineering and Decision Making* (13:4), 258–278.

Wallsten, T. S., Budescu, D. V., and Zwick, R. (1993). Comparing the calibration and coherence of numerical and verbal probability judgments. *Management Science* (39:2), 176–190.

Wallsten, T. S., Budescu, D. V., Zwick, R., and Kemp, S. M. (1993). Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bulletin of the Psychonomic Society* (31:2), 135–138.

Wan, Q., Hu, S., Zhang, Y., Wang, P., Wen, B., and Lu, Z. (2024). "It Felt Like Having a Second Mind": Investigating Human-AI Co-creativity in Prewriting with Large Language Models. *Proceedings of the ACM on Human-Computer Interaction* (8:CSCW1), 1–26.

Wang, K. and Nickerson, J. V. (2017). A literature review on individual creativity support systems. *Computers in Human Behavior* (74), 139–151.

Wang, W. and Benbasat, I. (2016). Empirical assessment of alternative designs for enhancing different types of trusting beliefs in online recommendation agents. *Journal of Management Information Systems* (33:3), 744–775.

Wang, X. and Yin, M. (2021). Are explanations helpful? A comparative study of the effects of explanations in ai-assisted decision-making. In *26th International Conference on Intelligent User Interfaces*, 318–328.

Warwick, K. and Shah, H. (2016). Can machines think? A report on Turing test experiments at the Royal Society. *Journal of experimental & Theoretical artificial Intelligence* (28:6), 989–1007.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511815478.

Wechsler, D. (1958). The measurement and appraisal of adult intelligence. *Academic Medicine* (33:9), 706.

Weiner, J. (1980). BLAH, a system which explains its reasoning. *Artificial Intelligence* (15:1-2), 19–48.

Wilder, B., Horvitz, E., and Kamar, E. (2020). Learning to complement humans. *arXiv preprint arXiv:2005.00582*.

Williams, K. D. and Karau, S. J. (1991). Social loafing and social compensation: The effects of expectations of co-worker performance. *Journal of Personality and Social Psychology* (61:4), 570.

Wintle, B. C., Fraser, H., Wills, B. C., Nicholson, A. E., and Fidler, F. (2019). Verbal probabilities: Very likely to be somewhat more confusing than numbers. *PLoS One* (14:4), e0213522. https://doi.org/10.1371/journal.pone.0213522.

Wolf, F. D. and Stock-Homburg, R. M. (2023). How and when can robots be team members? Three decades of research on human–robot teams. *Group & Organization Management* (48:6), 1666–1744.

Woodman, R. W., Sawyer, J. E., and Griffin, R. W. (1993). Toward a theory of organizational creativity. *Academy of Management Review* (18:2), 293–321.

Wright, J. L., Chen, J. Y., and Barnes, M. J. (2018). Human–automation interaction for multiple robot control: the effect of varying automation assistance and individual differences on operator performance. *Ergonomics* (61:8), 1033–1045.

Xu, Z. and Sheng, V. S. (2024). Detecting AI-Generated Code Assignments Using Perplexity of Large Language Models. In *Proceedings of the AAAI Conference on Artificial Intelligence* (38:21), 23155–23162.

Yam, K. C., Tan, T., Jackson, J. C., Shariff, A., and Gray, K. (2023). Cultural differences in people's reactions and applications of robots, algorithms, and artificial intelligence. *Management and Organization Review* (19:5), 859–875.

Yang, H.-L. and Tang, J.-H. (2004). Team structure and team performance in IS development: a social network perspective. *Information & Management* (41:3), 335–349.

Yang, Q., Steinfeld, A., Rosé, C., and Zimmerman, J. (2020). Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3313831.3376301.

Yeomans, M., Shah, A., Mullainathan, S., and Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making* (32:4), 403–414.

Yin, M., Wortman Vaughan, J., and Wallach, H. (2019). Understanding the effect of accuracy on trust in machine learning models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12.

Završnik, A. (2020). Criminal justice, artificial intelligence systems, and human rights. In *ERA forum* (20:4), 567–583.

Zedelius, C. M., Müller, B. C., and Schooler, J. W. (2017). *The science of lay theories. How Beliefs Shape Our Cognition, Behavior, and Health*. Springer. https://doi.org/10.1007/978-3-319-57306-9.

Zercher, D., Jussupow, E., Benke, I., and Heinzl, A. (2025). How Can Teams Benefit From AI Team Members? Exploring the Effect of Generative AI on Decision-Making Processes and Decision Quality in Team–AI Collaboration. *Journal of Organizational Behavior* (0:1). https://doi.org/10.1002/job.2898.

Zercher, D., Jussupow, E., and Heinzl, A. (2025). Team Climate in Team-AI Collaboration: Exploring the Role of Decisional Ownership and Perceived AI Team Membership. In *ECIS 2025 Proceedings*. URL: https://aisel.aisnet.org/ecis2025/human_ai/human_ai/8.

Zhan, Z., He, L., Tong, Y., Liang, X., Guo, S., and Lan, X. (2022). The effectiveness of gamification in programming education: Evidence from a meta-analysis. *Computers and Education: Artificial Intelligence* (3), 100096.

Zhang, Y., Liao, Q. V., and 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*, 295–305. https://doi.org/10.1145/3351095.3372852.

Zhou, E. and Lee, D. (2024). Generative artificial intelligence, human creativity, and art. *PNAS Nexus* (3:3), 52. https://doi.org/10.1093/pnasnexus/pgae052.

Zhu, M. and Zhang, M. (2017). Network Analysis of Conversation Data for Engineering Professional Skills Assessment. *ETS Research Report Series* (2017:1), 1–13. https://doi.org/10.1002/ets2.12191.