Int. J. Human–Computer Studies 205 (2025) 103622

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Uncovering the dynamics of human-AI hybrid performance: A qualitative meta-analysis of empirical studies

Dóra Göndöcs [a], Szabolcs Horváth [b], Viktor Dörfler [c,d,*]

[a] Széchenyi István University, Győr, Hungary
[b] Bellstreet Ltd, Budapest, Hungary
[c] University of Strathclyde Business School, Glasgow, United Kingdom
[d] Corvinus University of Budapest, Hungary

ARTICLE INFO

Keywords:
Human-AI collaboration
Hybrid performance
AI design
Qualitative meta-analysis
Performance factors
Interconnectedness
Literature review

ABSTRACT

Human-AI collaboration is an increasingly important area of research as AI systems are integrated into everyday workflows and moving beyond mere automation and augmentation to more collaborative roles. However, existing research often overlooks the dynamics and performance aspects of this interaction. Our study addresses this gap through a review of empirical AI studies from 2018–2024, focusing on the key factors influencing human-AI collaboration outcomes within the spectrum of Human-Centered Artificial Intelligence (HCAI).

We identify 24 critical performance factors that influence hybrid performance, grouped into four categories using thematic analysis. Then, we uncover and analyze the complex, non-linear interdependencies between these factors. We present these relationships in a factor dependency graph, highlighting the most influential nodes.

The graph and specific factor interactions supported by the papers reveal a quite complex web, an interconnectedness of factors. As opposed to being an easy-to-predict combination of inputs, human-AI collaboration in a given context likely leads to a dynamic, evolving system with often non-linear effects on its hybrid performance. Our findings and the previous research on automation technologies suggest that the application of AI tools in collaborative scenarios would benefit from a comprehensive performance framework. Our study intends to contribute to this future line of research with this initial framework.

1. Introduction

Artificial Intelligence (AI) has become a transformative force in various sectors, including healthcare, education, business, and leisure with numerous studies underlining this impact (Braun et al., 2022; Cabitza et al., 2023; Choi and Schwarcz, 2025; Dwivedi et al., 2021; Loske and Klumpp, 2021; Revilla et al., 2023; Vas, 2016).

This dynamic growth is largely driven by technological advances in machine learning (ML) and neural networks (Cabitza et al., 2023; Gilpin et al., 2018; Holzinger, 2018b; Jordan and Mitchell, 2015; LeCun et al., 2015; Richens et al., 2020) and most recently, the rise of agentic AI (Bornet et al., 2025).

Beyond automation, AI systems are increasingly in a collaborative mode (Haesevoets et al., 2021; Inkpen et al., 2023; Jain et al., 2022; Jarrahi, 2018; Xu et al., 2020)with humans in complex workflows: especially in decision-making tasks (Chen et al., 2023; Chong et al., 2022; Gomez et al., 2023a; Jakubik et al., 2022; Jarrahi, 2018; Lindebaum et al., 2020; Morrison et al., 2024; Richens et al., 2020; Schemmer et al., 2022; Wang and Yin, 2023; Weiss et al., 1978).

As a result, the role of AI is shifting from passive task augmentation (Acemoglu and Restrepo, 2018; Coombs et al., 2020; Davenport and Glover, 2018; Davenport and Kirby, 2015; Janssen et al., 2019; Jarrahi, 2018; Leyer and Schneider, 2021; Miller, 2018; Raisch and Krakowski, 2021; Vas, 2016) to interacting and even collaborating with human decision-makers, sometimes as peers (Blaurock et al., 2024; Bossen and Pine, 2023; Cabitza et al., 2021; Cabrera et al., 2023; Cao et al., 2023; Dell'Acqua, 2022; Dellermann et al., 2021; Fügener et al., 2021a; Gomez et al., 2023b; Lai et al., 2022; Kazuo Okamura and Seiji Yamada, 2020; Sowa et al., 2021; Tschandl et al., 2020; Wang et al., 2019). Although AI involvement in augmenting human capabilities is expanding, the

Declaration: The authors confirm that there are no relevant financial or non-financial competing interests to report.

This research was supported by the National Research, Development and Innovation Office (NKFIH) under grant number 2024-1.2.3-HU-RIZONT-2024-00030

* Corresponding author.
E-mail addresses: dora.gondocs@gmail.com (D. Göndöcs), szabolcs.horvath@bellstreet.me (S. Horváth), viktor.dorfler@strath.ac.uk (V. Dörfler).

https://doi.org/10.1016/j.ijhcs.2025.103622
Received 15 December 2024; Received in revised form 3 August 2025; Accepted 5 September 2025
Available online 6 September 2025

understanding of effective human-AI collaboration and hybrid performance, in our view, remains underdeveloped.

Despite significant advancements in AI and its growing use in practice, current AI literature lacks comprehensive performance frameworks that address the dynamic complexity of human-AI collaboration. Most research focuses on how AI can complement or enhance human tasks, keep emphasizing growth in efficiency and improved human performance. Furthermore, these studies often overlook how AI really interacts with and supports human users, and they do not explore features of effective human-AI collaboration, and its outcome, which we refer to as hybrid performance. We argue that AI should be put in a more collaborative role, raising critical questions about designing these systems to work together with humans enhancing, rather than diminishing, human agency and decision-making.

In order to understand human-AI collaboration better, we conduct a qualitative meta-analysis of empirical AI studies focusing on the following question: **"What are the most influential factors of human-AI hybrid performance?"** And by implication, we ask **"How should these factors be considered when designing and implementing AI in everyday workflows to foster human-AI collaboration?"**

Our goals are to identify the critical performance factors that influence the quality and efficiency of human-AI collaboration and hybrid performance and to map the interconnections of these factors and understand their nature better. By answering these questions, we contribute to the scholarly understanding as well as practice of AI systems design to support the development of hybrid performance management and, ultimately, a more effective human-AI systems in practice.

Our paper is structured as follows:

Section 2 presents the knowledge background and the knowledge gap in the literature.

Section 3 details the methodological framing of our study and the methods used for data analysis.

Section 4 introduces the paper selection and data collection process.

Section 5 presents the completed review and analysis and the findings.

Section 6 discusses the previously introduced findings in the light of the extant literature.

Section 7 highlights the paper's implications of the findings.

Section 8 concludes our exploration and indicates future research directions.

2. Knowledge background

To provide context for our study, we discuss how AI represents a new phase in automation, shifting from a tool-based approach to a more collaborative form of interaction. This shift aligns with the evolving Human-Centered Artificial Intelligence (HCAI) approach (Riedl, 2019; Shneiderman, 2020b), which emphasizes designing AI systems to enhance human capabilities while ensuring they are transparent, equitable, privacy-conscious, and under human control (Ozmen Garibay et al., 2023; Pisoni G, 2021; Shneiderman, 2020b). We then examine the key features of AI models that impact human-AI interaction, focusing on performance. We could not identify any models that effectively address the joint human-AI performance, and this gap provides scope for the contribution of this paper.

2.1. AI: a new chapter in automation

In line with the advances of AI developments, the traditional automation landscape has started shifting towards AI-based automation and augmentation to support human users' capabilities and performance (Davenport, 2018; Davenport and Glover, 2018; Davenport and Harris, 2017; Davenport and Kirby, 2015, 2016).

Unlike traditional automation, HCAI prioritizes a user-centric perspective, aiming to design systems that complement human strengths while addressing societal and ethical considerations (Nabizadeh Rafsanjani and Nabizadeh, 2023; Ozmen Garibay et al., 2023). According to this concept such human traits are critical to fostering meaningful collaboration, highlighting the importance of systems designed to augment and empower users rather than replace them (Herrmann, 2024; Yang et al., 2021).

AI-driven automation is used in many industries and in various job roles. It follows the typical pattern of traditional automation, where human workforce is replaced by machines, computers (Acemoglu and Restrepo, 2018; Davenport and Harris, 2017; Davenport et al., 2010; Davenport and Kirby, 2015, 2016). AI-based automation impact tasks rather than full jobs and is more relevant for white-collar workers (Davenport, 2018; Davenport and Harris, 2017). Though AI can replace certain tasks, most human skills remain difficult to automate, such as complex reasoning, judgment, analogy-based learning, abstract problem-solving, and a combination of physical activity, empathy, and communication skills.

Any kind of automation aims to determine the distribution or allocation of tasks between humans and machines (Dearden et al., 2000; Hollnagel and Bye, 2000; Janssen et al., 2019; Milewski and Lewis, 1997; Parasuraman et al., 2000; Press, 1971; Sheridan, 2000; Sheridan and Maschinenbauingenieur, 2002). The evolution of computer-based automation (Janssen et al., 2019; Parasuraman et al., 2000; Sheridan and Verplank, 1978) and the parallel development of AI (Buchanan et al., 1969; Feigenbaum, 1992; Minsky, 1988; Moravec, 1988; Newell and Simon, 1956; Peitgen et al., 1992; Russell and Norvig, 2020) set the scene for our exploration. While both concepts originated in the 1950s, they developed separately for many decades. The impact of computer-based automation, mainly driven by ERP systems, transformed the workplace, replacing paper-based administrative tasks with computer platforms (Janssen et al., 2019; Parasuraman and Mouloua, 2018; Parasuraman et al., 2000; Sheridan, 1992; Sheridan and Maschinenbauingenieur, 2002). At the same time, the introduction of the first expert systems marked the early development of AI (Feigenbaum, 1977; Russell and Norvig, 2020; Swartout et al., 1991).

The convergence of these two fields became clear as computer-based automation, including expert systems, changed the work of white-collar employees. The growing use of computers led to the development of larger software systems, and traditional software engineering opened the door for rule-based systems as a new phase of computer automation (Janssen et al., 2019; Lemaignan et al., 2017; Swartout et al., 1991). From a management decision-support view, human-computer automation follows strict rules, while AI-based automation, including expert systems using symbolic logic and machine learning (ML) in artificial neural networks (ANN), provides predictions that still require human judgment for final decisions (Dörfler, 2022; Feigenbaum, 1992; Gomez et al., 2023a; Göndöcs and Dörfler, 2024; Richens et al., 2020; Vaswani et al., 2017). The success of recent AI methods like deep learning (DL) (LeCun et al., 2015), convolutional neural networks (CNNs) (CNN, see LeCun et al., 1989; O'Shea and Nash, 2015), and generative AI (Gozalo-Brizuela and Garrido-Merchan, 2023; Vaswani et al., 2017) has sparked a new wave of automation in organizations (Choy et al., 2018; Davenport, 2023; LeCun et al., 2015; Montazeri et al., 2016; Tschandl et al., 2020). This wave reflects the principles of HCAI by emphasizing user adaptability, ethical design, and transparency in AI systems to improve interactions and outcomes (Riedl, 2019; Shneiderman, 2020c).

Coombs et al. (2020) introduce the concept of Intelligent Automation, which refers to the automation of knowledge-based tasks using AI. The authors explain that, beyond its technological benefits, this type of automation can also increase business value.

Dwivedi et al. (2021) evaluate the potential of AI from technological, organizational, and societal view in various industries. They point out that though AI can enhance and potentially replace humans in various tasks. However, the real challenge is how organizational leaders can use AI-enabled automation effectively to improve human performance.

Based on this development, we argue that human-computer and human-AI interactions share a similar foundation. While computer-

based automation aimed to replace human workers (Acemoglu and Restrepo, 2019; Bainbridge, 1983; Billings, 1997; Hoffman and Johnson, 2019; Kaber and Endsley, 1997; Wickens et al., 2015), human-AI interaction focuses more on supporting humans (Ebben, 2022; Leyer and Schneider, 2021; Miller, 2018; Raisch and Krakowski, 2021; Wilson and Daugherty, 2018). In augmentation, AI helps improve human performance, and this support could grow into a collaborative relationship (Jarrahi, 2018; Zahedi and Kambhampati, 2021).

The convergence of automation and AI is not only transforming the nature of work but also opened discussions on the 'future of work' (Acemoglu and Restrepo, 2019; Brynjolfsson and McAfee, 2014; Ford and Despeisse, 2016). Employees often resist automation because they fear it will threaten their job security and salary. However, their attitude toward augmentation is more positive. They see augmentation as support, where AI helps overcome human weaknesses and enhances their abilities.

Instead of being replaced, people can feel empowered by using AI in their work. AI can help them improve their skills and make better decisions (Raisch and Krakowski, 2021; Wilson and Daugherty, 2018). This partnership between humans and AI can create a more balanced work environment. AI can boost work quality, increase productivity, and lead to a more positive work experience (Zahedi and Kambhampati, 2021).

AI-driven automation should not only focus on efficiency, but it also impacts how humans and machines work together. AI's ability to handle complex tasks makes it a potential partner or peer colleague.

2.2. AI models' features and capabilities related to human-AI interaction

Any human-AI collaboration's success depends on the features and capabilities of the AI models used. AI models, particularly machine learning, including deep learning, operate based on probabilistic principles. That makes these models more complex and less predictable.

Recent advancements in AI, build on neural network architectures, emphasizing statistical (or probabilistic) methods over deterministic models (Wang et al., 2020). These approaches do not assume a completely well-structured context and enable the use of models like natural language processing and image recognition (Jarrahi, 2018; Tschandl et al., 2020; Willemink and Noël, 2019; Xu et al., 2014). However, such methods rely on vast datasets to predict outcomes, introducing variability that can lead to unpredictable or erroneous results (Daws, 2020; Dörfler, 2022; Guo and Hao, 2020; Heaven, 2019). The complexity of neural networks creates real challenges in understanding how these models operate. This opacity is often referred to as the "black box" problem, which makes it difficult to diagnose errors or biases (von Eschenbach, 2021; Gomez et al., 2023b).

AI models can perform outstandingly in narrow tasks that would require high levels of human expertise, but they can easily fail on simple problems that humans handle effortlessly (Moravec, 1988). AI models can achieve or even excel in adult-level or expert-level performance in certain areas. On the other hand, it can easily struggle with the tasks of a one-year-old child, such as perception and mobility. Current AI excels in narrow domains, such as diagnosing specific illnesses based on context-specific data (Acemoglu and Restrepo, 2019; Billings, 1997; Göndöcs and Dörfler, 2024; Tschandl et al., 2020), despite ambitions to develop Artificial General Intelligence (AGI).

The need for complementarity (Jarrahi, 2018) between humans and AI arises from the foundational differences between the human mind and AI. Machine learning models can learn from data autonomously, nevertheless, it still requires human supervision (Amershi et al., 2014; Choi et al., 2016; Gilpin et al., 2018; Organizing Committee of the Madrid 2017 Critical Care Datathon et al., 2019; Wang et al., 2020). Examples like counterfactual inference tasks and causal definitions (Richens et al., 2020; Rizzi, 1994; Weiss et al., 1978) illustrate this need for human oversight.

Interpretability and explainability often used as transparency are basic and crucial requirements for any AI model. It is critical to understand the focus or output of complex AI systems like image recognition (Bertrand et al., 2023; Gilpin et al., 2018; Herm et al., 2023; Holzinger et al., 2019; Lipton, 2018; Luo et al., 2019). Despite that interpretability can be achieved through visual cues and simplified proxies, AI remains a black box for its human users (Bertrand et al., 2023; Gilpin et al., 2018; Luo et al., 2019). Transparency means understanding how the outcome was achieved by AI, and through that provides users with insights into AI operations (Baryannis et al., 2019; Holzinger et al., 2019; Kim et al., 2023a). Göndöcs and Dörfler (2024) argue that in some professions, such as dermatology, explainability shall focus on understanding the scientific grounding of AI operation. Trust in AI requires confidence from the humans that AI models perform as expected. This outlines not using interpretability and explainability interchangeably, as explainable models are interpretable, but the reverse is not always true (Gilpin et al., 2018).

The relationship between explainability and responsibility is central to building a transparent human-AI relationship, leveraging their respective strengths in hybrid performance. The AI literature typically posits explainability as a prerequisite for trust, acknowledging that trust cannot be established without understanding how AI works (Gilpin et al., 2018; Lipton, 2018; Luo et al., 2019). Different levels of interpretation may be relevant to different people at varying mastery levels and professions (Baryannis et al., 2019; Göndöcs and Dörfler, 2024; Holzinger et al., 2019). Responsibility for AI's actions remains with humans for the foreseeable future (Coeckelbergh, 2020; Dennett, 1998; Floridi and Sanders, 2004).

We propose building a transparent relationship between humans and AI to leverage their strengths and complementarity and thus try to optimize the hybrid performance. Human users should have a clear picture of the used AI model's features and capabilities relating to the specific tasks and working contexts. The key is to balance human cognitive strengths and AI capabilities, implementing AI to complement and compensate for human cognitive constraints (Hoffman and Johnson, 2019; Shneiderman, 2020a, 2020b).

Interpretability can support the human user to better understand the AI-generated decisions, while explainability provides deeper insights into the processes that lead to certain decisions of AI. Finally, adaptability enables AI systems to adjust to human behaviors and preferences, and with that enhance the overall outcome.

In this section, we summarized the key features of AI models to ground our understanding of how these features affect the performance of AI, human users, and the combined human-AI hybrid performance. Next, we analyze the process of shifting from tool-based automation to human-AI collaboration. In line with this exploration, our main goal was to find a framework or model that explains how these features influence performance.

2.3. From automation to human-AI collaboration: the "Performance gap"

The key idea of (tool-based) automation is to achieve performance improvement by reducing the human component and intervention in tasks and processes (Acemoglu and Restrepo, 2018; Bainbridge, 1983; Billings, 1997; Dekker and Woods, 2002; K.Goldberg, 2012; Manzey et al., 2012; Parasuraman and Mouloua, 2018; Parasuraman et al., 2000; Sheridan, 1992). Automation research focuses on making systems more efficient, productive, and reliable. It deals with systems that can work on their own, usually in well-organized settings, for long periods. It also looks at how to set up these environments clearly and effectively (K. Goldberg, 2012; Parasuraman et al., 2000; Sheridan and Parasuraman, 2005).

Introducing AI systems do not differ much from traditional automation regarding their main goals of improving performance (Hoffman and Johnson, 2019; Kaber and Endsley, 1997; Kaber et al., 2000; Manzey et al., 2012; Parasuraman and Mouloua, 2018; Schreckenghost et al., 2014; Wickens et al., 2015). AI models are developed and trained to

produce at least the same but rather better results than the human performance before (Davenport and Harris, 2017; Davenport et al., 2010; Davenport and Kirby, 2015; Endsley, 2017; Kaber and Endsley, 1997; Parasuraman and Mouloua, 2018). Organizations also want to implement AI tools to achieve performance gains, such as efficiency improvement on cost-benefit basis, or any other quality type of features (Dell'Acqua et al., 2023; Ebben, 2022; Wang et al., 2020; Wuest et al., 2016).

Research on computer automation and human-computer interaction already identified several factors that are highly relevant to the overall performance of such systems (Billings, 1997; Kaber and Endsley, 1997; Parasuraman et al., 2000; Sheridan and Parasuraman, 2005; Sheridan and Maschinenbauingenieur, 2002).

On the other hand, recent AI systems are using more collaborative type of technologies with a focus on augmentation rather than full replacements as it was in traditional automation (Davenport and Glover, 2018; Jarrahi, 2018; Leyer and Schneider, 2021; Miller, 2018). Such AI systems can work together with humans to enhance their performance rather than simply replacing tasks (Davenport and Kirby, 2015; Miller, 2018; Raisch and Krakowski, 2021). Nevertheless, these new types of interactions come along with new challenges and highly complex issues. While AI research's main focus is to improve the models' performance, and thus they can excel in a lab setting, the same model faces critical issues in real-life applications (Gentile et al., 2023; Shin, 2021; Steyvers and Kumar, 2024; Wang et al., 2024; W. Wang et al., 2023). Many studies conclude that combining AI and human efforts doesn't necessarily lead to better outcomes or improved joined performance (Buçinca et al., 2021; Lemus et al., 2023; Lu et al., 2024; Sankaran et al., 2022; Schemmer et al., 2022; Weber et al., 2023; Wilson and Daugherty, 2018; Xu et al., 2020; Yildirim et al., 2023; Zhang et al., 2023). Such interaction is influenced by various factors, that also highlight that effective collaboration requires more than just high-performing AI (Cabitza et al., 2023; Sankaran et al., 2022).

Concluding from the above we believe that the proliferation of human-AI interaction and especially the collaborative use of AI in knowledge work and organizations in general will require a careful design and treatment regarding the expected performance outcomes. We propose that a proper theoretical background building on available empirical evidence might promote the development of this area.

We conducted a systematic search across various databases to find papers that have this collaborative way of interaction with a performance aspect in their focus. We were seeking to find any relevant performance framework or model in the field. Despite using different search terms and criteria, we did not find any papers that provide a comprehensive approach or offer a framework that helps understand and evaluate the human-AI hybrid performance. Nevertheless, we have found some promising evidence in some papers suggesting that the complexity of performance factors in human-AI collaboration is an important area to explore.

2.4. Clarifying key terms

To avoid confusion and ensure consistency, here we briefly clarify how we use the terms "*interaction*", "*augmentation*", and "*collaboration*" throughout the paper. We do not aim to provide formal or comprehensive definitions here, but rather to explain very briefly how we apply these terms in the context of our study.

We use "*interaction*" as the broadest term, meaning any kind of relationship or exchange between humans and AI. This includes all types of human-AI engagement, as for example, when AI is used as a tool or assistant, but it can also refer to real human-AI collaboration.

We use "*augmentation*" in the knowledge background section, where we describe the shift from automation (full replacement of humans) to augmentation, meaning that AI helps or supports human work instead of replacing it.

However, our main focus in this work is on "*collaboration*", the specific, collaborative type of interaction, where humans and AI cooperate with each other, like peers or teammates to achieve a shared goal.

3. Methodological framing

In this section, we outline the methodological framing of our paper, defining the scope, philosophical underpinnings, theoretical perspective, and the methods used to complete a thorough review work. We build on the systematic review approach to identify the relevant literature and we complete a method-based developmental literature evaluation (Petticrew, 2001; Templier and Paré, 2015).

3.1. Philosophical and theoretical positioning

This study has been conducted in a moderate interpretivist framing (Dörfler, 2023). Acknowledging the subjectivity, intentionality, judgement, pre-understandings of both the authors of the studies articles as well as of the researchers conducting this study, we can focus on the lived experience of AI implementation and are well positioned to deal with the methodological diversity of the analyzed studies.

Instead of adopting a theoretical lens, as any lens limits what can be seen, for this exploratory study we adopted a "lens-less" approach. We have considered three perspectives: Distributed Cognition Theory (Hutchins, 1995), Socio Technical Systems (Emery and Trist, 1965; Trist and Bamforth, 1951), and Activity Theory (Vygotsky, 1978, 1979). Each of these approaches offers significant benefits but each also limits what can be seen – as any lens does (Brickson, 2021). We have decided to adopt the so-called phenomenon-based theorizing instead, the basic idea of which is to allow the phenomenon under scrutiny to guide the theorizing process whichever way it goes (Bas et al., 2023; Fisher et al., 2021; Ployhart and Bartunek, 2019). Admittedly, this approach is not commonly used in meta-analysis, but as we have a well-defined phenomenon that we are interested in, AI implementations, and we aim at understanding this phenomenon, we believe that this was a meaningful choice (Langley, 2021). This approach also makes it easy to copy with the philosophical and methodological plurality of the included studies.

3.2. Study design

We achieve rigor in our meta-analysis, we employ explicit, systematic methods in the selection and inclusion of studies, and assessing their relevance via using meta-analysis to gain further insights from their findings (Liberati et al., 2009; Templier and Paré, 2015). Within that use qualitative meta-analysis (Timulak, 2009, 2013), a qualitative variant of meta-analysis (Petticrew, 2001; Templier and Paré, 2015) in which insights from primary studies serve as data for a secondary study aimed at increasing the scope and the depth in understanding the field (Hoon, 2013; Sandelowski et al., 1997; Timulak, 2009, 2013). We implement qualitative meta-analysis using thematic analysis (Liberati et al., 2009; Templier and Paré, 2015) and dependency graphs.

As main protocol for data collection, we use the PRISMA method (Page et al., 2021), the Preferred Reporting Items for Systematic Reviews and Meta-Analyses. By following the PRISMA guidelines, we implement a comprehensive and transparent process for identifying, screening and selecting relevant papers for our study.

We run systematic search across six online databases. Based on the first round of searches, we compiled an initial dataset of primary studies (Bartels, 2013; Meade and Richardson, 1997). Then we ran the screening of the papers in two stages: first, by screening papers based on their titles and abstracts, and second, by reviewing the full texts of papers that passed the initial screening.

To establish the scope of our study, in line with our research question, we carry out an exploratory search to determine the keywords and key terms as part of our search strategy. We define the specific inclusion and exclusion criteria for screening and selecting all the relevant papers in a systematic way. We decide to run our search for the period between

2018 and 2024, with this we are able to cover the overwhelming majority of papers employing recent (mostly deep learning and to a lesser extent large language models, LLM) AI models focusing on human-AI interaction.

Our data analysis consists of two phases. In the first phase, we coded the insights from the chosen papers qualitatively using an approach based on thematic analysis (Braun and Clarke, 2006, 2012, 2014, 2019). This method allows the use of a priori codes, derived from existing literature and prior knowledge in the field of AI, as well as in vivo codes that emerged during our analysis work. Besides this flexibility in coding, we chose thematic analysis for its effectiveness in meta-analysis (Aguinis et al., 2011; Point et al., 2017; Timulak, 2013), focusing on the insights from the papers rather than coding their full texts. The coding process is described in further detail in Section 5.1.

In the second phase of data analysis the identified factors were mapped and linked according to the interactions described in the analyzed studies, resulting in a graph. Examining the dynamics of the interactions between the factors we obtain important insights regarding the nonlinear nature of the system of factors.

By employing the above-described carefully designed search strategy along with the applied systematic methods during data analysis, we wanted to ensure the robustness of our study, making the meta-analysis comprehensive, in order to uncover the critical performance factors and these factors' interrelationships that influence the quality and efficiency of human-AI collaboration.

3.3. Common limitations of a qualitative meta-analysis

In this paper, we use a qualitative meta-analysis to review and interpret findings from several implementation studies. We are aware that this method has certain limitations, as described in earlier works. For example, researchers have pointed out that some of the depth, meaning, and context of the original studies can be lost during synthesis (Timulak, 2009, 2013).

Different studies often use different methods or theoretical perspectives, which can make it difficult to combine their results in a consistent way (Hoon, 2013; Sandelowski et al., 1997; Timulak, 2009, 2013). Also, because the analysis is based on already interpreted findings, there is a risk of adding further interpretation bias (Sandelowski et al., 1997; Timulak, 2013). Some studies may not report their methods clearly, making it hard to assess their quality or compare them (Liberati et al., 2009; Templier and Paré, 2015). Others have noted that qualitative meta-analyses sometimes focus too much on common findings and ignore less common or conflicting results, which can reduce the richness of the overall picture (Hoon, 2013; Sandelowski et al., 1997; Timulak, 2009, 2013). Ethical concerns may also arise, especially when the original studies did not mention possible reuse of the data (Liberati et al., 2009; Templier and Paré, 2015).

Despite these challenges, qualitative meta-analysis is a useful tool to bring together insights and highlight patterns that might not be visible in single studies and we used a suitable study design to cope with the philosophical and methodological diversity of the analyzed papers (Section 3.2) and we specifically explored the limitations of these papers (Section 4.5).

4. Selecting the empirical studies

We start our work by defining the specific focus of our work in line with the literature gap and research question from above. Our main focus is "empirical research on human-AI interaction focusing on collaborative use of AI tools, models typically in organizations, knowledge work settings and decision-making in general with direct or indirect references to the joint performance of the interaction".

In line with our focus, we set up inclusion and exclusion criteria. That was followed by the identification of relevant search terms and their respective formats for different databases. We then compiled specific search strings for each database.

We wanted to ensure that our approach is well-structured and rigorous. Since hybrid performance was the core focus of this review, we excluded studies that were purely technical or algorithmic in nature. Instead, we aimed to identify empirical AI studies that explicitly featured human-AI collaboration and could be analyzed from a performance management perspective.

4.1. Inclusion and exclusion criteria

These are the inclusion and exclusion criteria we defined.

Inclusion Criteria

- Emphasis on collaboration rather than automation, focusing on decision-making, knowledge work, and expert use of AI.
- Empirical papers, including lab experiments, user studies, real-life case studies, and interview-based studies.
- Completed research articles, with a preference for journal papers but also including conference papers and preprints from arXiv to capture recent studies.
- Inclusion of medical use papers due to significant examples of human-AI collaboration in this field.
- Date ranges from 2018–2024, covering recent, mostly applied deep learning methods in AI models and systems.

Exclusion Criteria

- Technical papers on machine learning techniques, model training, or development unless they address the interaction aspect.
- Papers on software design, user interface, or user experience unless focusing on performance outcomes.
- Studies related to robotics, manufacturing, or other physical automation, as our focus is on algorithmic decision-making.
- Papers on education, training, creativity, government, military, or public sectors unless they focus on AI collaboration.
- Studies on customer behavior, ethics, bias, or team dynamics, as these are outside our performance-focused scope.

4.2. Search terms and keywords applied

To gather the relevant literature, we performed an initial exploratory search to identify the keywords and refine our search strings. Our systematic search was completed across five databases, including Web of Science (WoS), Scopus, ACM Digital Library (ACM-DL), PubMed, and AISeL. We also added studies (excluding duplications) from Google Scholar using its less structured, free-form search engine, to ensure comprehensive coverage.

4.2.1. Search strings

In line with the defined keywords, we developed specific search strings based on format required from that database.

Web of Science (WoS): We searched for articles with titles containing "AI," "artificial intelligence," "machine learning," "large language model," "LLM," or "ChatGPT," and topics (titles, abstracts, or keywords) including "study," "studies," or "experiment," combined with "human-AI" or "AI-assisted." We excluded document types such as review articles and research areas like education, robotics, chemistry, materials science, environmental sciences, instrumentation, linguistics, and public health. This search yielded 944 results.

Scopus: We searched for articles with titles containing "AI," "artificial intelligence," "machine learning," "large language model," "LLM," or "ChatGPT," and title, abstract, or keywords including "study," "studies," or "experiment," combined with "human-AI interaction," "human-AI collaboration," "human-AI team," or "AI-assisted decision making." We excluded subject areas such as arts, physics, biochemistry, materials science, environmental science, energy, chemical engineering,

chemistry, nursing, and earth sciences. We made this search string more specific compared to what we used for WoS since it returned a large number of not relevant titles and many overlapping with WoS results. This search returned 716 results.

ACM Digital Library (ACM-DL): We searched for articles with titles containing "AI," "artificial intelligence," "machine learning," "large language model," "LLM," or "ChatGPT," and abstracts containing "study," "studies," or "experiment," combined with "human-AI interaction," "human-AI collaboration," "human-AI team," or "AI-assisted decision making." This search yielded 112 results.

PubMed: We searched for articles with titles containing "AI," "artificial intelligence," "machine learning," "large language model," "LLM," or "ChatGPT," and titles or abstracts containing "study," "studies," or "experiment," combined with "human-AI interaction," "human-AI collaboration," or "AI-assisted decision making." This search yielded 61 results.

AISeL: We searched for articles with "AI" in the title and "human-AI" and "experiment" or "study" in the abstract. This search yielded 97 results. Table 4.1 includes the specific search strings used and the results obtained from the specific databases.

4.2.2. Additional search in google scholar

Due to the lack of structured search options in Google Scholar, we used specific keywords to identify relevant studies through free-form searches.

The following keywords and terms were used: “human-AI collaboration”, “human-AI interaction”, "human-AI performance”, “human-AI hybrid performance”, “human-AI team”, “human-AI augmentation”, “AI-assisted decision-making” coupled with “empirical”, “experiment”, “study”/”studies”. A variation of “machine learning”, “LLM” and “ChatGPT” were also used.

Through this process and the snowballing (of review papers on empirical AI topics) referred to above, we manually screened and selected an additional **28** studies from Google Scholar, which were subsequently added to our final dataset.

Our systematic search across various databases yielded a total of 1958 relevant studies.

Table 4.1
Search strings used and results obtained.

Database	Search String	Results
Web of Science (WoS)	TI=("AI" OR "artificial intelligence" OR "machine learning" OR "large language model" OR "LLM" OR "ChatGPT") AND TS=("study" OR "studies" OR "experiment") AND TS=("human-AI" OR "AI-assisted")	944
Scopus	TITLE ("AI" OR "artificial intelligence" OR "machine learning" OR "large language model" OR "LLM" OR "ChatGPT") AND TITLE-ABS-KEY ("study" OR "studies" OR "experiment") AND TITLE-ABS-KEY ("human-AI interaction" OR "human-AI collaboration" OR "human-AI team" OR "AI-assisted decision making")	716
ACM Digital Library (ACM-DL)	Title:("AI" OR "artificial intelligence" OR "machine learning" OR "large language model" OR "LLM" OR "ChatGPT") AND Abstract:(("study" OR "studies" OR "experiment") AND ("human-AI interaction" OR "human-AI collaboration" OR "human-AI team" OR "AI-assisted decision making"))	112
PubMed	(("AI"[Title] OR "artificial intelligence"[Title] OR "machine learning"[Title] OR "large language model"[Title] OR "LLM"[Title] OR "ChatGPT"[Title]) AND ("study"[Title/Abstract] OR "studies"[Title/Abstract] OR "experiment"[Title/Abstract])) AND ("human-AI interaction"[Title/Abstract] OR "human-AI collaboration"[Title/Abstract] OR "AI-assisted decision making"[Title/Abstract])	61
AISeL	Title AND abstract:"human-AI" AND abstract	97

Source: Own research result.

4.3. Results of the data collection process

Here we summarize the results of our systematic data collection process completed in line with the protocol of the applied variant of meta-analysis, the PRISMA method (Petticrew, 2001).

Identification: First, we identified the relevant papers using predefined keywords and search terms across the selected databases. We developed specific search strings for each database in line with the required format. This search resulted in a total of 1930 relevant studies, as the initial dataset.

Removing Duplications: We then started refining this dataset by removing duplicates. This left us with 1469 papers.

Screening (Eligibility and Inclusion): After that, we conducted two rounds of screening. The first round involved screening the titles and abstracts of the selected papers, which drastically reduced the number of relevant papers from 1469 to 271. In the second round, we screened the full texts of these 271 papers applying our criteria leading to 178 items. By adding the 28 found through Google Scholar resulted in a total of 206 studies as the final dataset. Table 4.2 summarizes the results of our systematic search including the above-described phases.

4.3.1. Final dataset

The above-outlined rigorous search and screening process led us to the final set of 206 empirical papers. These papers collectively represent a wide range of industries, functions, situations, and applications. They cover healthcare (medical imaging, diagnosis, clinical decision support), finance, retail, logistics, legal, and creative industries among others. Several functional areas are involved like finance, sales, human resources, customer service, and engineering. The main themes of the papers reflect a diverse field like the variations of how humans and AI (potentially) work together, trust and reliance, explaining and interpreting how AI models operate, and others. Researchers used various methods that are quite relevant to our empirical criteria including experimental or quasi-experimental studies, field studies and case studies of real-world applications, user studies and interviews, survey-based studies and mixed methods, and a few simulations. For the first review, we find that these papers address our initial goal and criteria quite well and lend themselves to our thematic analysis that will follow.

4.4. Temporal trends in human-ai collaboration research

Building on the data collection described above, here we explain why we chose to focus on the period from 2018 to 2024 and what the

Table 4.2
Results of the Systematic Search with PRISMA Protocols.

Databases /No of Papers	identification	after duplications removed	screening (after title + abstract)	eligibility (after full text)
WOS	944	940	115	79
SCOPUS	716	460	128	81
ACM_DL	112	8	8	7
AIESEL	97	45	14	7
PUBMED	61	16	6	4
TOTAL No of Papers without Google Scholar	**1930**	**1469**	**271**	**178**
Google Scholar	28	28	28	28
TOTAL No of Papers with Google Scholar	**1958**	**1497**	**299**	**206**

Source: Own research result.

consequences of this choice are. We wanted to capture papers with modern AI methods, particularly DL models. These emerged a few years earlier but papers targeting human-AI collaboration studies based on DL models start to appear mostly from 2018. Earlier studies often relied on older, rule-based expert systems or traditional ANN models, which do not fully represent the collaboration challenges found in recent papers.

The period from 2018 marks a significant rise in research on AI-assisted decision making and collaborative scenarios. The research moves beyond automation towards interactive, explainable systems and human-AI workflow configurations.

Though LLMs became popular already in 2023, it takes a few years for collaboration research to reach maturity using these models. This aligns with our findings, in the 2018–2024 period we identified and included only 9 papers addressing LLMs. We find that collaboration with LLM models and LLM-based applications create similar challenges as specific, narrower DL systems. Some factors may be stronger or more pronounced with these LLM-based systems given their wide coverage of applications and conversational nature. Further to that, COVID-19 sped up digital work and supported the rise of LLMs, but research on remote AI collaboration using these models has only started to emerge from 2025.

There is a likelihood that additional factor(s) will be found given the growing body of research focusing on augmentation and human-AI collaboration with LLMs. This may be a good target scope for a review in a few years.

4.5. Common limitations in the reviewed studies

To get an impression of the strengths and limits of current research of human-AI collaboration, we analyzed the reported limitations in 20 randomly selected studies from our final dataset. This helped us take a more critical look at the quality of the evidence and the state of the field.

We found several common issues. Many studies focused on specific domains such as medicine, law, or customer service, which limits the scope of the results. Some used small or unbalanced samples such as mostly students or online workers, making more general insights problematic. Tasks were often too simple or unrealistic, and most studies only looked at short-term use. In many cases, the AI gave limited feedback or user control, making the validity findings about real collaborations questionable. Some studies also reported problems issues with overreliance on early AI suggestions, lack of user training, or poor explainability.

This reflection on the reported limitations shows that much of the research is still at an early stage. Our framework of interconnected factors can help bring more structure to the field and highlight areas for improvement, such as using more realistic tasks, studying long-term collaboration, and better supporting trust and understanding. A more detailed list of these limitations of the 20 studies is included in the appendix (Section 10.1).

5. Analysis and findings

As mentioned before, we used the systematic method of thematic analysis to analyze our final dataset of 206 papers. We have been carrying out the analysis in two phases. In the first phase, a thematic analysis was used to identify patterns from the findings of those papers. As a result of this analysis, we identified 24 factors that affect hybrid performance, that we collected in four groups.

In the second phase of our analysis, we analyzed the interaction of these factors using a dependency graph. In the next two subsections, we describe the analysis process and discuss our findings in detail.

This process was grounded in a detailed database that systematically assessed and analyzed the Title, Category, Focus, Methodology, and Key Findings of each paper. This database provided a structured foundation for our analysis, ensuring a comprehensive approach to uncovering insights.

5.1. The coding process: identifying and synthetizing the performance factors

The purpose of the coding process was to qualitatively interpret the analyzed studies, in order to identify emergent themes. Initially, we defined a list of a priori codes including terms like “human trust”, “AI performance” and “workflow configuration”, etc. As we progressed with the coding process, we refined these codes and identified new concepts, such as “mental models”, “delegation”, and “time pressure”, that came out from the data as in vivo codes. Through these, we could capture new, unique aspects of human-AI interaction.

During the coding process, we extracted 1–3 key findings (empirical results) from each study, particularly aiming at ones with performance-related insights (e.g., “explanations can help humans learn AI errors” or “trust in AI increases human intention to cooperate with AI”). These individual insights were used to form our third-level categories. As we encountered new insights, we continuously refined our codebook through overlapping codes and updating definitions to better reflect nuances in the literature.

To ensure a consistency and transparency of the performance factors, we created short descriptions for each factor and used a shared code structure across all papers. In order to increase coding reliability we used iterative team discussions, especially in cases of ambiguity or disagreement (cf bracketing through transpersonal reflexivity in Dörfler and Stierand, 2021; Olekanma et al., 2022).

This combination of a priori and in vivo codes enabled us to refine our understanding of the performance factors influencing human-AI collaboration. As a result of this thematic analysis, we identified 24 performance factors, that we grouped into four overarching categories reflecting the complex nature of human-AI collaboration:

- **AI Factors**: Capabilities, Performance, Transparency, Adaptability, and Interactivity.
- **Human Factors:** Performance, Self-Confidence, Trust, Reliance on AI, Expertise, Mental Models, Decision Authority, Cognitive Load, and Adaptability.
- **Interaction Factors:** Communication Mode, Complementarity, Roles Clarity, Workflow Configuration, and Delegation.
- **Task-Related Factors:** Task Category, Complexity, Time Pressure, Risk Level, and Other Related Processes.

This categorization, together with the consistent brief descriptions of each factor, contributes to the field by offering a structured and comprehensive framework to guide both research and practice in human-AI collaboration. (Table 5.1).

See the following table, which present the factor groups, the performance factors with a brief description.

These synthetized performance factors formed the third level of our coding hierarchy, representing aggregated insights from the literature. Through iterative refinement, we merged similar codes, adjusted definitions, and ensured that the final set of codes accurately captured the diversity and complexity of the factors influencing human-AI collaboration. This rigorous and comprehensive process synthesized empirical evidence into a robust framework that integrates both established concepts and emergent themes.

In the next step of our analysis, we wanted to see how the factors influenced each other in each study. We use a qualitative pairwise coding for that to capture the direction of each relationship based on the empirical findings.

5.2. Mapping the papers

We mapped each paper to the performance factors it addressed (Table 5.2). It is important to note that one paper could be linked to more than one factor, so the total number of papers is higher than the selected 206 papers. It is interesting to see how many papers were linked

Table 5.1
The 24 performance factors' brief description.

Factor	Factors' Description
I. AI Factors	
AI Capabilities	Refers to the range of functions, tasks, and operations that the AI system is designed to perform. It encompasses the specific skills and competencies of the AI, such as data processing, pattern recognition, predictive analytics, and decision support.
AI Performance	Includes the precision and consistency with which the AI system performs its tasks. Accuracy refers to the correctness of AI outputs, while reliability indicates how consistently the AI delivers accurate results over time.
AI Transparency	Involves the extent to which the AI system's decision-making processes are visible and understandable to users. It includes the ability of the AI to explain how it arrives at its decisions, making the system's operations more interpretable and trustworthy.
AI Adaptability	Refers to the AI system's ability to adjust to new data, changing environments, or user feedback. This includes the AI's capacity to learn from interactions and improve its performance over time.
AI Interactivity	Describes the quality and effectiveness of communication between the AI system and its users. It includes how responsive the AI is to user inputs, how well it can engage in a dialogue, and how it supports a dynamic exchange of information.
II. Human Factors	
Human Performance	Refers to the effectiveness, efficiency, and accuracy with which users carry out their tasks in collaboration with AI systems. It encompasses the ability to achieve desired outcomes, execute tasks promptly, and make accurate decisions.
Human Self-Confidence	Refers to the belief that users have in their own abilities to effectively perform tasks and make decisions when interacting with AI systems.
Human Trust	Refers to the level of confidence that users have in the AI system's capabilities and decisions. Trust affects how much users rely on the AI and how willing they are to integrate its outputs into their decision-making processes.
Human Reliance on AI	The degree to which a user depends on the AI system to perform tasks or make decisions. It's a behavioral outcome that is influenced by trust but also by other factors like task complexity, user expertise, and the AI's perceived capabilities.
Human Expertise	Refers to the knowledge, skills, and experience that the human users bring to their interactions with AI systems. It influences how effectively they can utilize the AI's capabilities and interpret its outputs.
Human Mental Models (Understanding of AI)	Are the internal representations that users form about how the AI system works, including its capabilities, limitations, and decision-making processes. Accurate mental models are essential for effective interaction with AI.
Human Decision Authority	Reflects the extent of control that human users retain over the decision-making process in collaboration with AI. It defines whether the AI is used as a decision-support tool or as a primary decision-maker.
Human Cognitive Load	Refers to the mental effort required to interact with the AI system. High cognitive load can impair decision-making, while well-designed AI systems can reduce cognitive load by simplifying tasks and presenting information clearly.
Human Adaptability	In human factors refers to the ability of users to adjust their strategies and behavior in response to the AI system's outputs and capabilities. It includes the flexibility to integrate AI recommendations and to work effectively in changing environments.
III. Interaction Factors	
Communication Mode	Refers to the channels and methods used for interaction between humans and AI, such as text, voice, visual interfaces or other specific modes of interaction (eg. prompting). The mode of communication can impact the clarity and effectiveness of the interaction.
Complementarity	Describes the synergy achieved when the strengths of humans and AI are combined to achieve better outcomes than either could alone. It focuses on how well the human and AI components of the system enhance each other's performance.
Roles Clarity	Involves the clear definition of responsibilities and tasks between human users and the AI system. Clear role definitions ensure that both parties know what is expected of them, reducing confusion and improving collaboration efficiency.
Workflow Configuration	Refers to the structure and sequence of tasks and decision-making processes within the collaboration. This includes whether the workflow is human-first, AI-first, or a hybrid approach, and how tasks are allocated between human and AI.
Delegation	In the context of human-AI interaction refers to the process by which tasks or decisions are assigned to the AI system by the human users. Effective delegation involves determining which tasks are best handled by AI and which require human judgment.
IV. Task Factors	
Task Category	Task category refers to the type of task being performed, such as decision-making, data analysis, or creative work. Different task categories may require different levels of AI involvement and human-AI interaction.
Task Complexity	Task complexity describes the level of difficulty and intricacy of the tasks involved in human-AI collaboration. Complex tasks often require more sophisticated AI capabilities and a higher degree of collaboration between humans and AI.
Time Pressure	Time pressure refers to the urgency with which tasks need to be completed. High time pressure can influence decision-making processes and may lead to greater reliance on AI to expedite tasks.
Risk Level	Risk level refers to the potential consequences and severity of outcomes associated with the task or decision-making situation in human-AI collaboration. Higher risk levels involve greater stakes, such as the possibility of significant financial loss, harm to individuals, or critical failures, which may necessitate more cautious decision-making, increased human oversight, and a careful evaluation of AI recommendations.
Other Related Processes	This factor includes any additional processes or systems that interact with or influence the human-AI collaboration, such as organizational workflows, regulatory compliance, and existing technology infrastructures.

Source: results of data analysis.

to each factor, as this shows how much attention each one has received in research. To make this easier to grasp, we grouped the factors into three categories based on how often they appeared in the studies. This helps highlight which areas have been studied more and which ones are still less explored.

Most studied factors (mentioned in more than 30 papers): *AI Transparency (79 papers), AI Performance (42), Human Trust (47), Human Reliance on AI (79), Human Mental Models (39)*

Moderately studied factors (mentioned in 10 to 30 papers): *AI Capabilities (10), AI Adaptability (11), AI Interactivity (11), Human Performance (20), Human Self-Confidence (18), Human Expertise (22),*

Table 5.2
The 4 factor groups, 24 performance factors with related papers.

Factors	Relevant Papers
I. AI Factors	
AI Capabilities (10 papers)	(Fügener et al., 2021b; Okamura and S. Yamada, 2020) (Bansal et al., 2019b; Cvetkovic and Bittner, 2022; Dell'Acqua et al., 2023; Eini-Porat et al., 2022; Flathmann et al., 2023b; Grisold and Schneider, 2023; Hsieh and Lee, 2024; Lebovitz et al., 2022; Pataranutaporn et al., 2023; Schreibelmayr et al., 2023)
AI Performance (42 papers)	(Andre et al., 2023; Bansal et al., 2021a; Baryannis et al., 2019; Bienefeld et al., 2023; Bossen and Pine, 2023; Braun et al., 2022; Bui et al., 2023; Cecil et al., 2024; Choi and Schwarcz, 2025; Chong et al., 2024, 2023, 2022; De Brito Duarte, 2023; Dell'Acqua, 2022; Dell'Acqua et al., 2025; Dhuliawala et al., 2023; Eini-Porat et al., 2022; Erlei et al., 2024; Fügener et al., 2021a, 2021b; Gu et al., 2024; Inkpen et al., 2023; Kahr et al., 2023; Kahr et al., 2024a and 2024b; Le et al., 2022; Lemus et al., 2023; Maier et al., 2022; Nascimento et al., 2023; Pandya et al., 2019; Peng et al., 2022; Pinski et al., 2023; Rechkemmer and Yin, 2022; Schaap et al., 2024; Tejeda et al., 2022; Tschandl et al., 2020; D.-Y. Wang et al., 2023; Wang et al., 2019; G. Wang et al., 2023; W. Wang et al., 2023; Xu et al., 2023; Zhu et al., 2024)
AI Transparency (79 papers)	(Abdel-Karim et al., 2023; Altintas et al., 2023; Bansal et al., 2019a, 2021b; Baryannis et al., 2019; Berberian et al., 2023; Blaurock et al., 2024; Bossen and Pine, 2023; Braun et al., 2023, 2022; Bussone et al., 2015; Cabitza et al., 2023, 2024; Cabrera et al., 2023; Cai et al., 2019; Calisto et al., 2021; Cao et al., 2024; Cassenti et al., 2023; Cecil et al., 2024; Chen et al., 2023; Chiang and Yin, 2021, 2022; Corvelo Benz and Rodriguez, 2023; De Brito Duarte, 2023; van der Zander et al., 2024; Ebermann et al., 2023; Fügener et al., 2021b; Fukuchi and Yamada, 2023; Gentile et al., 2023; Gu et al., 2021; Guo et al., 2024; Harbarth et al., 2024; Hauptman et al., 2024; He et al., 2023; Holstein et al., 2023; Hüllmann et al., 2023; Jakubik et al., 2022; Jiang et al., 2022; Kahr et al., 2023; Kahr et al., 2024a; Karran et al., 2022; Kim et al., 2024, 2023a; Lebovitz et al., 2022; Lee et al., 2024; Lee and Chew, 2023; Leichtmann et al., 2023a, 2023b; Lemus et al., 2023; Liu, 2021; Liu et al., 2021; Lu and Yin, 2021; Ma et al., 2023; Morrison et al., 2024; Mozannar et al., 2023; Mucha et al., 2021; Papenkordt, 2024; Park et al., 2022; Pataranutaporn et al., 2023; Robbemond et al., 2022; M. Schemmer et al., 2023b; Schreibelmayr et al., 2023; Shen et al., 2021; Shin, 2021; Sivaraman et al., 2023; Spillner et al., 2023; Subramanian et al., 2023; Suffian, 2024; Taesiri et al., 2022; Taudien et al., 2022a, 2022b; Tejeda Lemus et al., 2022; Tschandl et al., 2020; Vasconcelos et al., 2023; Walter et al., 2023; Wang and Ding, 2024; Wang and Yin, 2022, 2023; Y. Wang et al., 2024; Westphal et al., 2023; Wu et al., 2022; Zhang et al., 2023, 2020)
AI Adaptability (11 papers)	(Bansal et al., 2021b; Bhattacharya et al., 2024; Fügener et al., 2021b; Gao et al., 2021; Grisold and Schneider, 2023; Guzdial et al., 2019; Lu et al., 2024; Mozannar and Sontag, 2020; Nascimento et al., 2023; Okamura and Yamada, 2020; Salikutluk et al., 2024)
AI Interactivity (11 papers)	(Buçinca et al., 2021; Cheng et al., 2024; Danry et al., 2023; Gomez et al., 2023b; Gu et al., 2021; Lee et al., 2021; Liu et al., 2021, 2023; Muijlwijk et al., 2024; Papenkordt, 2024; Wu et al., 2022)
II. Human Factors	
Human Performance (20 papers)	(Beese and Fahse, 2023; Bondi et al., 2022; Bossen and Pine, 2023; Braun et al., 2023, 2022; Cao et al., 2024; Dang et al., 2023; Danry et al., 2023; Dell'Acqua, 2022; Fügener et al., 2021b; Gentile et al., 2023; Hemmer et al., 2023; Kim et al., 2024; Lee et al., 2024; Loske and Klumpp, 2021; Rastogi et al., 2022; Revilla et al., 2023; Tschandl et al., 2020; Walter et al., 2023; Y. Wang et al., 2024)
Human Self-Confidence (18 papers)	(Allen and Choudhury, 2021; Bondi et al., 2022; Chong et al., 2024, 2023, 2022; Fogliato et al., 2022; Fügener et al., 2021a; He et al., 2023; Holstein et al., 2023; Lu and Yin, 2021; Ma et al., 2024; Niraula et al., 2024; Papachristos et al., 2021; Schaap et al., 2024; Subramanian et al., 2023; Tejeda et al., 2022; Tschandl et al., 2020; Wang et al., 2022)
Human Trust (47 papers)	(Altintas et al., 2023; Bansal et al., 2021b; Berberian et al., 2023; Bui et al., 2023; Cao et al., 2023; Cecil et al., 2024; Chong et al., 2024, 2023; Corvelo Benz and Rodriguez, 2023; Cvetkovic and Bittner, 2022; De Brito Duarte, 2023; Dhuliawala et al., 2023; Gentile et al., 2023; Glienke et al., 2023; Hauptman et al., 2024; Hou et al., 2023; Hsieh and Lee, 2024; Inkpen et al., 2023; Kahr et al., 2023; Kahr et al., 2024a and 2024b; Karran et al., 2022; Kim et al., 2023a, 2023b; Kuang et al., 2024; Leichtmann et al., 2023b; Li et al., 2023; Liu, 2021; Liu et al., 2021; Ma et al., 2023; Maier et al., 2022; Muijlwijk et al., 2024; Niraula et al., 2024; Okamura and S. Yamada, 2020; Park et al., 2022; Pataranutaporn et al., 2023; Schreibelmayr et al., 2023; Shin, 2021; Sivaraman et al., 2023; Spillner et al., 2023; Taudien et al., 2022a, 2022b; Wang et al., 2019; Wang and Ding, 2024; W. W. Wang et al., 2023; Wang and Yin, 2023; Westphal et al., 2023; Yao et al., 2024; Zhang et al., 2023, 2020)
Human Reliance on AI (79 papers)	(Abdel-Karim et al., 2023; Adam et al., 2024; Allen and Choudhury, 2021; Altintas et al., 2023; Ardichvili et al., 2024; Bansal et al., 2021a, 2021b; Blaurock et al., 2024; Bondi et al., 2022; Bossen and Pine, 2023; Buçinca et al., 2021; Bussone et al., 2015; Cabitza et al., 2024; Calisto et al., 2021; Cao et al., 2023, 2024; Cecil et al., 2024; Cheng et al., 2024; Chiang et al., 2024; Chiang and Yin, 2021, 2022; Chong et al., 2023, 2022; De Brito Duarte, 2023; van der Zander et al., 2024; Ebermann et al., 2023; Eini-Porat et al., 2022; Fahse and Schmitt, 2023; Flathmann et al., 2023b; Fogliato et al., 2022; Fukuchi and Yamada, 2023; Glienke et al., 2023; Gomez et al., 2023b; Guggemos, 2024; Guo et al., 2024; Harbarth et al., 2024; He et al., 2023; Herm et al., 2023; Holstein et al., 2023; Hou et al., 2023; Hüllmann et al., 2023; Jakubik et al., 2022; Jiang et al., 2022; Jussupow et al., 2021; Kahr et al., 2024a; Kelly et al., 2023; Kim et al., 2024, 2023b; Lee and Chew, 2023; Lemus et al., 2023; Li et al., 2022, 2023; Lu et al., 2024; Lu and Yin, 2021; Ma et al., 2024; Morrison et al., 2023, 2024; Mozannar et al., 2023; Mucha et al., 2021; Niraula et al., 2024; Pataranutaporn et al., 2023; Pinski et al., 2023; Rastogi et al., 2022; Reverberi et al., 2022; Robbemond et al., 2022; Salikutluk et al., 2023; Salimzadeh and Gadiraju, 2024; Salimzadeh et al., 2024; Schemmer et al., 2023a; Schemmer et al., 2023b; Schreibelmayr et al., 2023; Shaikh and Cruz, 2023; Sivaraman et al., 2023; Sowa et al., 2021; Subramanian et al., 2023; Swaroop et al., 2024; Taesiri et al., 2022; Tejeda et al., 2022; Tejeda Lemus et al., 2022; Vasconcelos et al., 2023; D.-Y. Wang et al., 2023; G. Wang et al., 2023; Westphal et al., 2023; Yao et al., 2024; Zhang et al., 2022; Zhu et al., 2024)
Human Expertise (22 papers)	(Allen and Choudhury, 2021; Ardichvili et al., 2024; Bulten et al., 2021; Cao et al., 2024; Choi and Schwarcz, 2025; Dell'Acqua et al., 2023; Herm et al., 2023; Holstein et al., 2023; Inkpen et al., 2023; Jiang et al., 2022; Jussupow et al., 2021; Kim et al., 2023b; Morrison et al., 2024; Ou et al., 2023; Qian and Wexler, 2024; Rondonotti et al., 2023; Tejeda Lemus et al., 2022; Thomas et al., 2024; Tschandl et al., 2020; W. Wang et al., 2023; Wang and Yin, 2022; Zhang et al., 2020)
Human Mental Models (39 papers)	(Allen and Choudhury, 2021; Bansal et al., 2021a, 2019a, 2019b; Blaurock et al., 2024; Buçinca et al., 2021; Cabrera et al., 2023; Cai et al., 2019; Cassenti et al., 2023; Chiang and Yin, 2021; Corvelo Benz and Rodriguez, 2023; De Brito Duarte, 2023; Dhuliawala et al., 2023; Fügener et al., 2021b; Gu et al., 2024; Guggemos, 2024; Harbarth et al., 2024; Herm et al., 2023; Holstein et al., 2023; Ibrahim et al., 2023; Jiang et al., 2022; Jussupow et al., 2021; Kelly et al., 2023; Lebovitz et al., 2022; Lee et al., 2021; Leichtmann et al., 2023b; Liu, 2021; Liu et al., 2023; Pataranutaporn et al., 2023; Pinski et al., 2023; Reverberi et al., 2022;

(*continued on next page*)

Table 5.2 (*continued*)

Factors	Relevant Papers
	Schemmer et al., 2023a; Shen et al., 2021; Shin, 2021; Suffian, 2024; Tejeda Lemus et al., 2022; C. Wang et al., 2024; Wang et al., 2019, 2022; Westphal et al., 2023)
Human Decision Authority (6 papers)	(Abdel-Karim et al., 2023; Adam et al., 2024; Berberian et al., 2023; Kong et al., 2024; Schaap et al., 2024; Westphal et al., 2023)
Human Cognitive Load (3 papers)	(Calisto et al., 2021; Eini-Porat et al., 2022; Jiang et al., 2022)
Human Adaptability (3 papers)	(Beese and Fahse, 2023; Guzdial et al., 2019; Sivaraman et al., 2023)
III. Interaction Factors	
Communication Mode (7 papers)	(Ardichvili et al., 2024; Choi and Schwarcz, 2025; Dang et al., 2023; Glauben et al., 2024; Kim et al., 2024; Lebovitz et al., 2022; Rapp et al., 2023)
Complementarity (17 papers)	(Bansal et al., 2021a; Brehmer et al., 2022; Chen et al., 2023; Fügener et al., 2021b; Hekler et al., 2019; Inkpen et al., 2023; Morrison et al., 2024; Revilla et al., 2023; Richens et al., 2020; Salikutluk et al., 2023; Tejeda Lemus et al., 2022; Wang et al., 2016; W. Wang et al., 2023; Wilder et al., 2021; Xu et al., 2023; Zhang et al., 2022; Zhang et al., 2020)
Roles Clarity (1 paper)	(Papachristos et al., 2021)
Workflow Configuration (21 papers)	(Abdel-Karim et al., 2023; Brehmer et al., 2022; Cabitza et al., 2023, 2021; Chiang et al., 2024; Fogliato et al., 2022; Gomez et al., 2023b; Gu et al., 2023; Henkenjohann and Trenz, 2024; Hu et al., 2024; Kuang et al., 2024; Lauer and Wieland, 2021; Li et al., 2022; Lobo et al., 2024; Lu et al., 2024; Sowa et al., 2021; Swaroop et al., 2024; D.-Y. Wang et al., 2023; Wu et al., 2022; Xu et al., 2020; Zhang et al., 2024)
Delegation (12 papers)	(Adam et al., 2024; Bondi et al., 2022; Cvetkovic and Bittner, 2022; Erlei et al., 2024; Fügener et al., 2021a, 2021c; Gao et al., 2021; Grisold and Schneider, 2023; Hemmer et al., 2023; Lai et al., 2022; Pinski et al., 2023; Taudien et al., 2022a)
IV. Task Factors	
Task Category (2 papers)	(Cao et al., 2023; Revilla et al., 2023)
Task Complexity (6 papers)	(Choi and Schwarcz, 2025; Chong et al., 2023; Kim et al., 2023b; Salimzadeh and Gadiraju, 2024; Salimzadeh et al., 2024; Westphal et al., 2023)
Time Pressure (6 papers)	(Cao et al., 2023; Gu et al., 2021; Lebovitz et al., 2022; Rastogi et al., 2022; Shaikh and Cruz, 2023; Swaroop et al., 2024)
Risk Level (5 papers)	(Cao et al., 2024; De Brito Duarte, 2023; Gu et al., 2021; Kim et al., 2023b; Li et al., 2023)
Other Related Processes (3 papers)	(Guo et al., 2024; von Richthofen et al., 2022; Ruissalo, 2024)

Source: Results of Data Analysis.

Complementarity (17), Workflow Configuration (21), Delegation (12)

Less studied factors (mentioned in fewer than 10 papers): *Human Decision Authority (6), Human Cognitive Load (3), Human Adaptability (3), Communication Mode (7), Roles Clarity (1), Task Category (2), Task Complexity (6), Time Pressure (6), Risk Level (5), Other Related Processes (3)*

This also shows that some factors, such as transparency and reliance, have been significantly studied, while others, such as decision authority or task-related aspects, have received much less attention.

Why are some factors studied less? There could be several reasons. Human–AI collaboration is still a developing field, so many topics have not been fully explored yet. Some factors might also be harder to study, depend on specific situations, or have just been missed by researchers so far. But simply because a factor is mentioned less often does not mean it is not important. For example, human decision authority, cognitive load, or task complexity might still play a key role in certain situations.

5.3. Interconnectedness emerge: graphing factors relationships

As presented in the table above we identified 24 relevant performance factors from the papers in our review. We noticed that many studies not only present findings related to one or a few of these factors but also provided evidence of the interaction between two or more factors. Thus we realized that our three-level coding structure may not fully capture the dynamic, interconnected nature of the factors. More specifically, key findings of these papers often describe how one factor influences another, showing that performance is shaped by relationships, not just isolated elements. In addition to these 24 factors, we also included the concept of "hybrid performance" as an outcome, since several factors have a direct impact on it.

To uncover these relationships, we reviewed the key findings of each selected paper, seeking directional links in the form "*x* affects *y*" or "x influences y" where x and y are among the 24 performance factors. Then using qualitative pairwise coding (not statistical), we analysed each paper's description of human-AI collaboration and coded pairs of factors accordingly. This process resulted in a total of 326 unique directional links between the performance factors.

An example of such pair is Human Trust → Human Reliance on AI, indicates that at least one paper provided evidence that human trust (a belief) influences how humans rely on AI (a behaviour). This relationship implies that the degree of trust users develop towards an AI system affects how they engage with it in practice, including how often they use it and under what conditions. Building upon the 326 identified directional pairs we created a dependency graph (see Fig. 5.1) where each arrow represents an empirically observed influence from one factors to another (Lin et al., 2013). This graph helps visualize the complex web of interdependencies between factors within a specific context and this also suggests that hybrid performance is shaped in complex ways, through chains of influence. Instead of isolated cause-effect links, hybrid performance emerges through a dense network of influences between interconnected factors.

The strength of the interdependencies is not measured through statistical weighting, but rather indicated by the number of studies discussing each connection as noted in Section 5.2. Each directional link in the graph represents a relationship empirically observed in at least one study. If a particular link appears in multiple studies, it suggests that the relationship is more frequently recognized in the literature and possibly also more relevant in real-life practice. However, we want to reinforce that the frequency of mention does not necessarily reflect true causal strength; it may also be shaped by current research focus.

By analyzing the graph, we identify the most connected factors (nodes) by counting the number of links attached to each. These central factors may have a stronger influence on the overall performance of human-AI collaboration. Even relationships supported by fewer studies may still be highly relevant and should not be overlooked.

A recurring example of such interdependence is the path from *AI Transparency* to *Human Trust*, to *Human Reliance on AI*, and finally to *Hybrid Performance*. Studies consistently show that when AI systems provide understandable, transparent outputs, users develop greater trust. This leads to more appropriate reliance on the system, which ultimately improves collaborative outcomes.

The dependency graph can also serve as a practical tool for AI system design. By mapping the critical factors in a given context and examining how they influence each other, designers can better target improvements and anticipate indirect effects. The graph highlights key points where changes in one factor may positively affect others. Continuous follow-up is also important: by tracking how these factors evolve during use, designers can make informed adjustments to better meet user needs and improve overall performance.

5.3.1. Key insights from the graph

The dependency graph visually shows the dynamic interplay among the 24 performance factors during any human-AI collaboration.

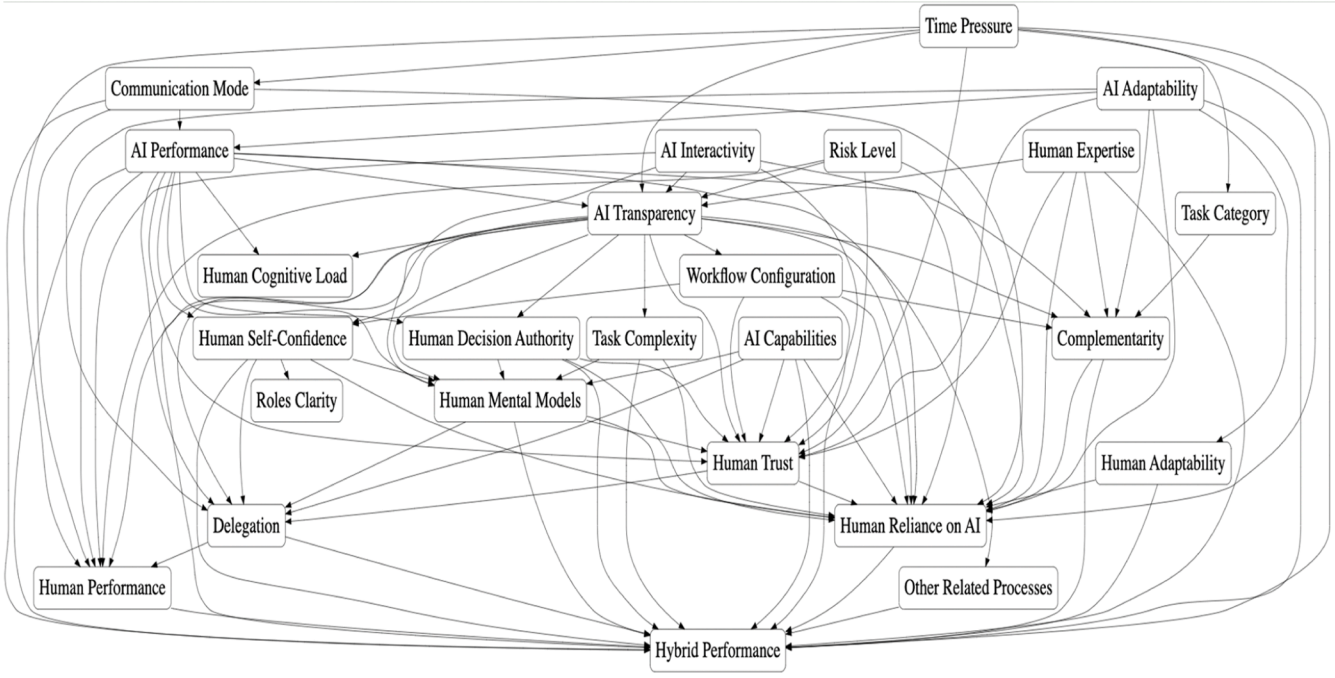

Fig. 5.1. The graph of the factor interconnections.
Source: Own research result.

To sum up the main insights, we found the following:

- **Dynamic Complexity:** The factors rarely work alone. For instance, one factor can interact with several others, and this way shaping hybrid performance in complex ways. For example: *Human Trust in AI* can influence not only *Reliance on AI*, but also *Decision-making Authority*, and *Cognitive Load*.
- **Non-Linear Relationships**: Some factors affect each other in non-linear ways. This means that even small changes in one factor can lead to much bigger, unexpected effects in others.
- **Hybrid Performance**: Many of these factors directly or indirectly, shape hybrid performance, as the outcome of how humans and AI work together. This highlights its central role in human-AI collaboration

Finally, we highlight that a factor dependency graph can be a useful tool in practice. It helps understand specific human-AI collaboration in a specific context by showing how the relevant factors are connected and influence each other. This makes it easier to find the crucial factors, see how changes in one can impact others, and plan improvements. The graph can also help track which factors need to be followed up over time and adjusted if needed. In practice, this approach can be made even stronger by adding statistical data to support decisions. Overall, we suggest that such visualization tool can give helpful insights during AI system design.

6. Discussion of review results

In this section, we discuss and analyze the findings from our meta-analysis. First, we look at each of the identified performance factors to understand their importance. In the second part, we explore how the interconnectedness of these factors plays out supported by examples from a number of papers.

By doing so, we aim to address how these insights can guide the development of AI systems that not only achieve technical excellence but also align with human needs, aspirations, and ethical considerations. This interconnected perspective not only advances our understanding of hybrid performance but also provides actionable insights into designing AI systems that effectively complement human capabilities.

In the following sections, we to unpack these themes in detail, that aim to enhance human-AI collaboration while addressing the challenges and opportunities identified in the literature.

6.1. The 24 performance factors

In this section we introduce the factors according to the four emergent groups; one subsection corresponding to each group.

6.1.1. AI factors

One of the most obvious factors when applying AI tools to support human work is the **capabilities of the AI system**. This will define the scope and potentially the utility of the AI and the way how it may be used by the human user. An example can be found in this paper (Eini-Porat et al., 2022) where clinicians desire more reliable and actionable predictions from AI, tailored to their workflow and tasks. But already the presentation of the capabilities of an AI system directly influences human perception, trust and adoption (Flathmann et al., 2023b; Hsieh and Lee, 2024). Built-in AI capabilities like learning (Grisold and Schneider, 2023) or reasoning about human mental attitudes are shown to have an effect on the collaboration. There may be tension between updating an AI system capabilities and maintaining compatibility with users' previous experiences, mental models and expectations (Bansal et al., 2019b). It was found that when updates were not in line with these even an improved accuracy model could harm joint performance.

The **AI performance** has traditionally been viewed as the primary metric for optimizing performance of human-AI collaboration. An AI model is typically trained on a constrained dataset tailored to a well-defined prediction task such as classifying a symptom, recommending a pass/fail decision in a quality inspection, or generating a new sentence based on a text input. Model training is performed for a specific outcome and accuracy measures. However, given its probabilistic way of working, model outcomes are rarely perfect, they come with a margin of error. These results in false positives and false negatives (e.g. healthy patients classified with a disease or missed disease cases in a diagnostic situation), inherent errors of machine learning models. Machine learning engineers use a number of techniques to find a good balance of

fitting the model to the available data and generalizing to new, unseen data and contexts. This trade-off can have significant consequences when the model is deployed in real-world applications, where its performance may vary under different conditions. Other trade-offs will also be present like accuracy versus interpretability especially in applications requiring transparent operations. Depending on the model choice (or other factors) some tools may be of higher accuracy, but human users will have more difficulty to understand their inner workings and interpret outcomes. Beyond accuracy consistency and reliability too are key performance metrics. Consistency refers to the model's ability to produce similar results across similar inputs while reliability ensures that the model performs predictably under varying input data and other conditions and over time. Experience shows that even models with human level or superhuman performance by design will not guarantee the same level when integrated in real-world workflows or paired with human users.

Within the context of human-AI collaboration our intuition may be that pairing a human user with a highly performant AI tool will surely lead to a better performance. There are a number of studies showing an improved human decision-making in collaboration with high performance AI tools (Bossen and Pine, 2023; Dell'Acqua et al., 2023; Kahr et al., 2024a; Pandya et al., 2019; Peng et al., 2022; Wang et al., 2019; Xu et al., 2023). On the other hand, there is also evidence that in certain situations a lower quality model will contribute to a better team performance (Dell'Acqua, 2022; Pandya et al., 2019) by improving human engagement or complementarity.

AI performance, especially accuracy has a widely studied direct influence over human trust and adoption or reliance on using the AI (Dhuliawala et al., 2023; Kahr et al., 2023; Kahr et al., 2024a and 2024b; Maier et al., 2022; Park et al., 2022; Rechkemmer and Yin, 2022). High model accuracy builds human trust and intention to use but AI errors may degrade this trust and the interaction (Kahr et al., 2023; Maier et al., 2022). An increased level of human trust likely makes them rely more on the AI. Depending on the actual task and the model performance this may improve or worsen the hybrid performance. AI errors have an impact through actual use when users rely on its faulty outcomes, accepting an AI advice when they should not (Cecil et al., 2024; Dhuliawala et al., 2023; Erlei et al., 2024; Kahr et al., 2023; Kahr et al., 2024b).

According to the definition that we adopted the **AI transparency** groups a number of key features of AI systems that relates to how humans are able to understand the workings of the system. It consists of explanations of outcomes, descriptions of system performance, confidence (typically displayed case-level), error boundaries, edge cases and expected behavior in certain situations, and guidelines. Transparency of AI systems, models are a widely covered research area, this is the factor in our study which occurred in the largest number of papers. The reasons for this may be two-fold: 1) this directly relates to the "black-box" nature (as mentioned above) of deep learning models overwhelmingly used in AI solutions 2) explainable AI (XAI) has been studied extensively from various angles including by researchers specializing in ML engineering.

There are various types of transparency discussed in the studies, here we summarize the most relevant:

- Feature importances: highlight the most salient features of the data influencing the AI outcomes
- Model confidences: these are confidence scores (like 0–100) or probability distributions indicating the level of certainty of the AI predictions
- Counterfactual explanations: help users understand model behavior under changing inputs (e.g. "what-if" scenarios)
- Model explanations: descriptions of training data and its distribution, internal workings of AI models, details of and reasoning about outputs
- Tutorials, instructions, onboarding: these serve to educate users about how to engage with the system effectively

AI transparency influences human trust and human reliance on AI use. More detailed AI explanations and confidence scores can have a positive impact on trust and reliance (Berberian et al., 2023; Braun et al., 2022; Cabitza et al., 2023; De Brito Duarte, 2023; Fügener et al., 2021b; Liu, 2021; Ma et al., 2023; Taudien et al., 2022a; Walter et al., 2023; Zhang et al., 2020). But there are a number of papers which found that the impact is not as straightforward or even negative (Altintas et al., 2023; Bansal et al., 2021b; Bussone et al., 2015; Cabitza et al., 2023; Cecil et al., 2024; Ebermann et al., 2023; Harbarth et al.; Hauptman et al., 2024; Holstein et al., 2023; Jiang et al., 2022; Lee and Chew, 2023; Liu, 2021). The authors of study (Hauptman et al., 2024) suggest that the high-explainability solution caused communication overload which was counter-productive. This paper (2022) finds that providing alternative advice and showing prediction confidence scores may hinder users' advice acceptance especially for users lacking proper knowledge. Transparent systems with well-explained outcomes even if wrong may be more "convincing" or reduce human vigilance and lead to over-reliance. Trust in AI is also affected by explanation type (De Brito Duarte, 2023; Lee and Chew, 2023; Leichtmann et al., 2023b; Ma et al., 2023; Taesiri et al., 2022; Wang and Ding, 2024).

AI explanations can contribute to better human task performance (Braun et al., 2022; Gentile et al., 2023; Lee et al., 2024; Tschandl et al., 2020; Walter et al., 2023). On the other hand hard to understand, imperfect or "bogus" explanations may mislead users including domain experts even if the AI is correct (Morrison et al., 2024; Spillner et al., 2023). Transparent AI systems help build human mental models while they interact with the system. Case level explanations or confidence scores have a direct, local effect on decision making in a particular instance. Repeated interactions and use of an AI system combine experiences and feedback that gradually builds users' mental models of system behavior and understanding of other crucial features (Bansal et al., 2019a; Cai et al., 2019; Fügener et al., 2021b; Shen et al., 2021; Tejeda Lemus et al., 2022). These human mental models will have a significant role in shaping human-AI collaboration. A few studies also suggest that presentation and design matters, since transparency may also increase cognitive load for humans and increase the decision time needed (or otherwise worsen the performance) (Calisto et al., 2021; Jiang et al., 2022; Westphal et al., 2023). And finally, people tend to use confidence scores or other similar AI model uncertainty information differently based on their skills or domain knowledge (Holstein et al., 2023; Lemus et al., 2023; Subramanian et al., 2023).

AI adaptability refers to the capacity of the AI system to adjust to changing environments, tasks or user needs. This ability may increase the effectiveness of the human-AI interaction by adapting to new conditions and learn from feedback. Personalization is an important form of AI adaptability. It highlights the need to make the interactions individual and achieve complementary performance by taking into account the human decision-makers abilities (Fügener et al., 2021b; Gao et al., 2021; Guzdial et al., 2019). An AI system may dynamically adjust its role. Team performance improved when the AI agent adapted its level of autonomy to the situation (Salikutluk et al., 2024). In another study the system defers decisions to the human only when needed depending on the human expertise showing flexibility in task delegation (Mozannar and Sontag, 2020). Another area of AI adaptation is trust calibration. (Okamura and S. Yamada, 2020) demonstrated that "adaptively presenting a TCC (trust calibration cue) could promote trust calibration more effectively than a traditional system transparency approach". Explanations and alternative recommendations ("second opinion") may also be adaptable to reduce over-reliance contributing to properly calibrated trust levels (Bansal et al., 2021b; Lu et al., 2024). Learning from feedback and experience allows systems to gradually refine their workflow, delegation mechanism and overall performance (Bhattacharya et al., 2024; Gao et al., 2021; Grisold and Schneider, 2023; Mozannar and Sontag, 2020; Nascimento et al., 2023).

Interactive AI systems allow dynamic and iterative exchange between a user and the system. **AI interactivity** can increase system

transparency and controllability as shown in (Wu et al., 2022) by introducing a chaining approach in an LLM model. (Papenkordt, 2024) finds "if users proactively demand an explanation of the recommendation, it becomes increasingly likely that they cognitively evaluate it". Implementing novel interface designs ("interactive explanations") enable users to manipulate input features and observe how AI outcomes change leading to a better understanding of the prediction (Liu et al., 2021). The study (Gu et al., 2021) introduces interactivity by developing and testing a tool that engages users in an active and iterative process in a medical diagnosis context. This includes iteratively accepting/rejecting image regions highlighted by AI, labeling on-the-fly to provide feedback to the system and adjusting pre-filled diagnosis dialogues. Human mental models are another area where interactivity plays a role. In a code generating LLM study (Liu et al., 2023) users are provided with natural language feedback which improves their understanding of the system's capabilities and limitations. Cognitive forcing interventions which make people engage more thoughtfully with the AI significantly reduced over-reliance (Buçinca et al., 2021).

6.1.2. Human factors

The study of **human trust** has gained a significant attention similarly to AI transparency since these are seemingly closely related in studies. The definition we use for human trust includes human confidence in AI for practical reasons (treated similarly in most papers). Another closely related factor that is often studied in tandem and (casually) used interchangeably with trust is human reliance on AI. We decided to keep reliance as a separate factor since trust is a belief while reliance is a behavior with all the pertaining implications in empirical studies. Trust influences reliance (or intention to rely) on AI, higher positively, lower negatively (Glienke et al., 2023; Hou et al., 2023; Li et al., 2023; Niraula et al., 2024; Schreibelmayr et al., 2023). While another study highlights the emotional component of trust as a predictor of interaction (Yao et al., 2024). Whereas trust is not a fixed "quantity", it evolves over time driven by interaction (Li et al., 2023).

As mentioned above **human reliance on AI** is a factor closely studied with trust. Even though used interchangeably in some studies most research distinguishes these two in measuring and explaining results. Human reliance may be affected by trust both in terms of the degree users rely on the AI outputs and the effectiveness of the interaction. "Trust calibration" refers to the appropriateness of trust-reliance relationship. Over-reliance may signal the use of AI advice even when it is not adequate while under-reliance misses opportunities to improve human performance by accepting AI recommendations. Human reliance is typically an outcome in the papers we analyzed ie. a dependent variable. The majority of the factors that we identified has some influence on the reliance to use the AI. A few examples for such factors: AI Transparency, Human Mental Models (Understanding of AI), Human Self-Confidence, Human Expertise, Human Trust, Workflow Configuration.

There are a number of studies with empirical support that find **human mental models** that shapes human user's understanding of AI, directly influencing human-AI collaboration. It is quite challenging to directly examine human mental models, the research methods and approaches used in the studied papers are primarily focusing on revealing its impact. The most often mentioned and analyzed relationship is through reliance ie. human mental models affect the use of AI tools. Human mental models shaped by other elements such as trust, cognitive abilities, learning and experience] (Allen and Choudhury, 2021; Bansal et al., 2021a; Buçinca et al., 2021; Guggemos, 2024; Herm et al., 2023; Jiang et al., 2022; Jussupow et al., 2021; Kelly et al., 2023; Patar-anutaporn et al., 2023; Reverberi et al., 2022; Schemmer et al., 2023a) has a consequence on how and when users rely on AI advice. Trust may be directly influenced by users understanding of AI (Dhuliawala et al., 2023; Wang et al., 2019).

Human self-confidence mostly affects reliance, or in other words the use of AI (Allen and Choudhury, 2021; Bondi et al., 2022; Chong et al., 2023, 2022; He et al., 2023; Lu and Yin, 2021; Ma et al., 2024; Niraula et al., 2024; Tejeda et al., 2022). Humans who are less confident in their own judgement are more likely accept an AI advice (Bondi et al., 2022; Tejeda et al., 2022), especially if the AI model is highly confident (Tejeda et al., 2022). While humans with high confidence tend more to reject incorrect AI advice (Chong et al., 2023, 2022). High self-confidence may lead to suboptimal use of AI due to algorithmic aversion, bias and overall under-reliance on AI (Allen and Choudhury, 2021; Lu and Yin, 2021; Niraula et al., 2024) This effect is also moderated by the influence of level of agreement between own assessment and the AI's (Lu and Yin, 2021; Niraula et al., 2024). Humans even has a tendency (also a widely studied cognitive bias) to overestimate their own competence (He et al., 2023). Otherwise, self-confidence may be a crucial component of good decision-making: accepting when being confident in their own judgement and being able to reject even if less confident thus avoiding over-reliance on faulty advice (Chong et al., 2023, 2022). Somewhat unexpectedly self-confidence may be low despite high skill level so even highly skilled individuals might fall into over-relying AI (Chong et al., 2022).

Considering **human expertise** studies indicate that AI typically improves performance for individuals with lower skill levels (Bulten et al., 2021; Dell'Acqua et al., 2023; Inkpen et al., 2023; Ou et al., 2023; Qian and Wexler, 2024; Thomas et al., 2024; Tschandl et al., 2020). However domain experts tend to trust AI less (Allen and Choudhury, 2021; W. Wang et al., 2023), as their knowledge and experiences leads them to question the validity of AI advice. While this skepticism (otherwise known as algorithmic aversion) may protect them from over-reliance in certain situations (Inkpen et al., 2023; Kim et al., 2023b; Morrison et al., 2024; Rondonotti et al., 2023; W. Wang et al., 2023) it may as well be a limiting factor in achieving better performance. Complementary performance is likely to improve when the human party has more extensive domain knowledge or task experience (W. Wang et al., 2023; Zhang et al., 2020).

Human decision authority, or the degree of control humans have when interacting with AI, impacts trust, understanding and reliance (Kong et al., 2024; Westphal et al., 2023), via perceived autonomy and ability to adjust system recommendations according to the findings. Users of AI systems often expressed their preference for a supportive AI tool stimulating reflective thinking and collaboration (Abdel-Karim et al., 2023; Adam et al., 2024; Kong et al., 2024)

Human adaptation refers to how human users of AI adjust their interaction strategies, techniques and behavior over time working with AI. Studies show users selectively applying various behavior patterns during interaction and behavior changes over time (Beese and Fahse, 2023; Sivaraman et al., 2023). These papers found specific adaptations that have been observed for other factors, like aligning mental models, calibrating trust, developing AI skills.

While **human performance** is most often measured as part of the overall hybrid performance of the human-AI collaboration, a few papers directly refer to its role (Beese and Fahse, 2023; Loske and Klumpp, 2021; Revilla et al., 2023). However, the majority of studies reflect on how other factors like AI Transparency, AI Performance, AI interactivity, Delegation are affected by human performance (Bondi et al., 2022; Braun et al., 2023; Cao et al., 2024; Fügener et al., 2021b; Gentile et al., 2023; Lee et al., 2024; Y. Wang et al., 2024)

AI Transparency may be seen as something useful though it may inadvertently cause **human cognitive load** issues (Calisto et al., 2021; Eini-Porat et al., 2022; Jiang et al., 2022). Providing explanations, more detailed model, output information or uncertainty information can increase complexity and processing time for the human involved (Westphal et al., 2023).

6.1.3. Interaction factors

Communication and interaction modes refer to channels, methods and techniques between human and AI, including techniques like prompting or providing natural language feedback. These modes like

interaction techniques (Fügener et al.) and behavioral strategies (Ardichvili et al., 2024; Choi and Schwarcz, 2025; Dang et al., 2023; Glauben et al., 2024; Kim et al., 2024; Rapp et al., 2023) directly affect how well users interact with AI tools and their overall performance.

The concept of **complementarity**, when humans and AI systems working together outperform either one alone, is a widely accepted principle in human-AI collaboration research. This performance improvement may be the consequence of not fully overlapping strengths/weaknesses when applied to a specific decision scenario. Many studies focus on the hybrid performance without explicitly measuring complementarity effects, but a few papers directly address this factor (Hekler et al., 2019; Salikutluk et al., 2023; Tejeda Lemus et al., 2022; Wang et al., 2016; Wilder et al., 2021; Zhang et al., 2022). For instance, algorithms that complement human weaknesses in medical diagnosis examples can enhance the hybrid performance as shown in. This paper (Zhang et al., 2022) finds that participants were able to assess the varying level of complementarity and adjusted their reliance accordingly.

Roles Clarity is a factor relatively underrepresented in our sample of papers compared to the others. We still opted to keep it because this concept may prove to be useful in situations where the task environment is not well defined than in most of the studies in our review. One noteable example can be found in (Papachristos et al., 2021) where users develop their own perception of the AI role during the collaboration and adjust their behavior to this role.

Workflow configuration plays a critical role in the effectiveness of the human-AI interaction by directly triggering the way how and when humans use the AI (Gu et al., 2023; Hu et al., 2024; Lauer and Wieland, 2021; Lobo et al., 2024; Sowa et al., 2021; Swaroop et al., 2024; Wu et al., 2022; Xu et al., 2020). Papers explore basic workflow designs such as the timing of AI use (first, second, parallel) or collaboration models (complementarity, substitutability, and modularity) (Xu et al., 2020) and their impact on the hybrid performance. Configurations that promote human engagement, critical thinking and reflective practices can positively influence performance (Abdel-Karim et al., 2023; Chiang et al., 2024; Wu et al., 2022; Zhang et al., 2024), while some papers revisit the impact on performance through reliance (Chiang et al., 2024; Fogliato et al.; Gomez et al., 2023b; Li et al., 2022; Lu et al., 2024; Sowa et al., 2021). Others find that specific configurations of the interaction like suppressing AI under certain conditions, giving alternative opinions (Sowa et al.), prior provisional response may contribute to appropriate reliance (Fogliato et al.; Lu et al., 2024; D.-Y. Wang et al., 2023). Interestingly, even the simplest choices such as AI-before vs AI-after configuration may impact reliance, accuracy and speed (Cabitza et al., 2023; Swaroop et al., 2024). Although one-step workflows may lead to anchoring bias and a consequent performance degradation (Fogliato et al.).

Delegation mechanisms in human-AI systems, including how tasks are allocated to either the human or the AI agent and the effectiveness of delegation behavior have a direct impact on hybrid performance (Adam et al., 2024; Bondi et al., 2022; Fügener et al., 2021a, 2021c; Lai et al., 2022). A few studies experiment with the effect of which party is delegating (human vs. AI) and observe performance differences (Hemmer et al., 2023)especially when there is uncertainty (Fügener et al., 2021a). Another interesting research avenue is conditional delegation which allows humans to set rules for when AI decisions are trustworthy (Lai et al., 2022).

6.1.4. Task factors

Studies suggest that the kind of task that is the subject of human-AI collaboration, the **task category** as we define it, matters when it comes to hybrid performance (Cao et al., 2023; Choi and Schwarcz, 2025; Inkpen et al., 2023; Revilla et al., 2023; Salimzadeh et al., 2024). Examples: type of forecasting, spatial vs. logical reasoning, prognostic vs diagnostic tasks in medical setting, task-specific tuning, open-ended vs. closed task structure.

Task complexity has a varied impact on other factors: may decrease overall performance (Chong et al., 2023; Salimzadeh and Gadiraju, 2024; Salimzadeh et al., 2024), studies (Kim et al., 2023b; Salimzadeh et al., 2024; Vasconcelos et al., 2023)found that increasing task difficulty leads to over-reliance, when tasks are more complex users' understanding of how to collaborate with AI becomes less accurate (Bansal et al., 2019a)

Time pressure can fundamentally alter the ways how users interact with AI systems influencing the frequency and scope of use and also the performance outcomes (Cao et al., 2023; Rastogi et al., 2022; Shaikh and Cruz, 2023; Swaroop et al., 2024). In some studies users rely more on AI when under time constraints (Shaikh and Cruz, 2023; Swaroop et al., 2024) but this may be limited by the type of task or task category (Holzinger, 2018a) also finds that certain forms of AI assistance are more beneficial under time-pressure compared to no-time-pressure situations. More time allocated to the process can improve the reliance when users are more critically assess AI advice (Cao et al., 2023; Rastogi et al., 2022). AI transparency may have limited benefits in a time-constrained environment like in medical settings (Gu et al., 2021; Lebovitz et al., 2022)where transparency has a capacity to improve understanding and quicker processing by humans. On the other hand, the added cognitive effort may easily negatively impact decision-making speed.

Risk level is mostly associated in the studies with AI transparency, trust and reliance (Cao et al., 2024; De Brito Duarte, 2023; Kim et al., 2023b; Li et al., 2023). Users are generally more critical in high-risk scenarios indicating less trust and intention to use.

Finally, adoption of AI tools in organizational workflows may require **other related processes** and task to be adjusted having an impact on the overall performance of the AI use (von Richthofen et al., 2022; Ruissalo, 2024).

6.2. Understanding the non-Linear effects in human-AI collaboration

The studies we cover in this analysis overall suggest that human-AI collaboration and its joint performance is not the product of simple additive relationships among contributing factors. Instead we observe non-linear effects, where small changes in some factors (e.g. AI model accuracy) cause large, disproportional shifts in others (e.g. trust in AI output). These may lead to unexpected, sometimes counterintuitive results.

In this context, non-linear effects refer to interactions among performance factors where the combined impact is not proportional to the contributing individual influences, or factor contribution leads to sudden shift in factor relationship.

Examples:

- Overreliance on highly accurate AI: while higher AI model accuracy tends to increase overall performance studies show that it may paradoxically reduce user vigilance. Users may accept AI recommendations uncritically thus leading to worse performance as compared to standalone. Some papers even show that lower accuracy models result in better overall performance. Thus, performance gain from AI model accuracy is not linear.
- Transparency and cognitive overload: by increasing the transparency of the AI model output, e.g. providing explanations, the human user's understanding will improve. A better understanding presumably leads to better judgement and collaborative performance. Adding explanations is not a clear-cut exercise – there may be inadequate representation, too much/unnecessary details or cognitive overload in a time critical situation. While some may be helpful, inadequate explanations can end up in worsening performance.
- Trust collapses on AI errors: even a small number of AI mistakes can cause a sharp decline in user trust and reliance, despite the AI's overall accuracy. This illustrates a tipping-point effect, where trust erosion is abrupt and not proportional to error frequency.

The non-linearities are important in understanding factor interactions, and also when designing and optimizing real-world human-AI systems. These challenge general assumptions and the possibility of understanding the factors in isolation. Effective system design should anticipate trade-offs, thresholds and tipping-points. We also believe that future research may address these situations better based on our study.

We provide specific examples of the above relationships and references to individual studies in the next two sections.

6.2.1. Interdependencies between the factors and their impact on performance

Taking into account the most connected factors and the key observations supported collectively by the analyzed papers, we highlight several typical interactions among various factors that can be found across multiple studies.

These interdependencies suggest that performance factors contribute to a complex, evolving environment, often leading to non-linear and sometimes surprising outcomes:

AI Transparency and Mental Models: The transparency of the AI system shapes human mental models about how the system work and the outcomes it generates. Human understanding, in turn, defines the level of trust the user develops in the AI. However, this trust may (or may not) lead to reliance on AI recommendations. When this reliance is appropriate, then it leads to optimal hybrid performance. In many cases trust is not well "calibrated" or not correlated positively with reliance, so it can result in sub-optimal joint performance and prevent humans from achieving effective complementarity with AI. Human trust and self-confidence also interact with each other and may have a significant role in the effectiveness of the collaboration.

AI Performance and User Perception: Users' perception of AI performance also determines their trust and confidence in the AI's capabilities and usefulness. This trust may be a deciding factor when and how users will use the AI. Contrary to what one might expect, not always the highest performing AI model that leads to the best team performance.

Transparency and Cognitive Load: In various contexts we found that AI transparency not only improves (potentially) human understanding but can cause higher cognitive load for the user. Explanations, confidence scores and similar information can inadvertently increase task complexity, and this effect is often moderated by the user expertise, cognitive abilities and domain knowledge. In high-stakes scenarios under time pressure, this increased complexity could negatively impact user trust and reliance on the AI.

Moderating Transparency: The optimal level of AI transparency is notably moderated by the AI model performance and shows that finding a good balance may be context dependent (on nature of task, expertise level, etc.). A lower performing AI system could provide more transparency (explanations, confidence levels) to help users in finding faulty AI suggestions. However, offering explanations has been shown to interact with AI uncertainty information (confidence score) as perceived by the human user. However, AI performance also affect human self-confidence and the human confidence in AI in unexpected and opposite ways.

Workflow Configuration: In terms of workflow configuration, the human-first vs. AI-first models have an impact on reliance and directly on hybrid performance. Under time pressure some configurations work better than others. Workflows that promote reflective, interrogative human practices lead to better human performance. Interactive workflows may positively influence human trust and human understanding (mental models) of the AI by helping reasoning about results, verifying correctness, providing feedback.

In the following section we demonstrate the dynamic, often non-linear interconnectedness of various factors using examples from selected papers. These empirical observations and evidence suggest that several factors are at play and their relationships and dependencies may lead to counterintuitive results.

6.2.2. Complex and non-linear interactions – specific examples

This section provides examples to show how different performance factors interact in human-AI collaboration. These examples help reveal some surprising patterns and challenge common assumptions about how humans and AI work together.

AI Performance and Human-AI Collaboration: AI performance is thought to directly contribute to the overall performance via its accuracy and speed. Although, as we noted earlier, in a collaborative scenario the hybrid performance will also depend on the human performance component and on how the human decision-maker rely on the AI tool. Still, one might expect that a higher performing AI is always better in any situation. But it is not the case as shown in (Dell'Acqua, 2022) that a lower quality AI may lead to better outcomes, depending on the effort, learning, and skillset of the humans involved. The authors (Dell'Acqua, 2022) note that "subjects with higher quality AI were less accurate in their assessments of job applications than subjects with lower quality AI.. these effects were driven by more experienced recruiters". From a different angle, a study (Bansal et al., 2021a) argues that AI models tuned to teamwork are better than high accuracy models. "Our experiments with linear and non-linear models on real-world, high-stakes datasets show that the most accuracy AI may not lead to highest team performance and show the benefit of modeling teamwork during training through improvements in expected team utility across datasets, considering parameters such as human skill and the cost of mistakes." (Bansal et al., 2021a)

AI Performance and Human Trust: AI performance has also an impact on human trust, just like as on human self-confidence. A high AI accuracy may increase the trust of its user (Kahr et al., 2023; Kahr et al., 2024a; Rechkemmer and Yin, 2022) but this also heavily depends on the reliability of the AI (Dhuliawala et al., 2023; Kahr et al., 2024b). A lower accuracy but more consistent AI tool can be more trustworthy. At the same time both human trust and self-confidence affect human reliance on using AI while also likely interacting with each other. The study (Chong et al., 2022) found that good decision-makers tend to accept AI advice when their self-confidence is high and were able to reject more AI advice than poor decision-makers even if they were not so confident in their own judgement. Poor decision-makers on the other hand fell into a vicious cycle when lowered self-confidence after accepting faulty AI suggestions led them to rely more on AI thus compounding poor performance (Chong et al., 2022). Another study (Chong et al., 2024) finds as well that users feel personally accountable and less self-confident when the AI performs poorly. Human trust in AI drops with poor AI performance but does not improve with good performance according to (Chong et al., 2023). This implies a non-linear relationship where negative events have a stronger weight than positive ones. In the same paper human self-confidence is shown to decrease with AI accuracy changes despite the direction of the change (Chong et al., 2023). Furthermore, confidence in AI and self-confidence affect decisions to accept AI input in opposite directions: "Surprisingly, the more confident designers are in the AI's ability, the less likely they are to accept AI suggestions, and the more confident designers are in their own ability, the more likely they are to accept AI suggestions." (Chong et al., 2023). While "The fact that showing confidence improved trust and trust calibration but failed to improve the AI-assisted accuracy is puzzling… even though showing confidence encouraged participants to trust the AI more in high-confidence zones, the number of trials where the human's predictions were better substitutes for AI's predictions in low-confidence zones was not enough to improve overall accuracy." (Zhang et al., 2020)

AI Transparency and Human User Impact: AI transparency may be able to improve human-AI collaboration, but explanations differ in their usefulness according to the skill, expertise or analytical capacities of the human users (Wang and Ding, 2024). Although visual explanations improved human decision-making in an experiment this did not translate to higher trust in the AI system or stronger intentions to use it (Leichtmann et al., 2023a). Explanations increased perceived task complexity which impaired users' trust and understanding, meanwhile

moderated by cognitive ability with higher ability leading to better outcomes (Westphal et al., 2023). ” Participants interacting with a transparent system differentiated between correct and incorrect recommendations less consequently than participants interacting with a non-transparent system. Thus, in a transparent system, users seem to follow incorrect recommendations more often indicating complacency.” (Harbarth et al.)

Trust and Human Reliance: Regarding trust and human reliance, we should not assume that higher self-reported trust in AI would correlate with relying more on it; for instance, (Wang and Ding, 2024) shows that users might follow AI recommendations even if not confident about them.

Time Pressure and Task Complexity: Time pressure alters human behavior in a way that it affects the whole human-AI decision-making process. Time constraints not only impact human performance but also make users to weigh between speed and accuracy (speed-accuracy trade-off) (Swaroop et al., 2024). The lack of time may trigger users to rely more on AI. Another study shows (Cao et al., 2023) that giving more time for users to assess AI advice makes them more likely to accept it.

Human Expertise and Non-Linear Dynamics: Human expertise (seniority level, domain knowledge) shows various non-linear patterns. Those with less expertise typically found to gain more from AI interaction. This study (Allen and Choudhury, 2021) found the opposite though: “low experience participants reject algorithmic advice primarily due to lack of ability to assess and use algorithmic recommendations”. More domain knowledge may be beneficial in collaboration with AI from various angles but can also lead to less than optimal reliance, especially when confident in their initial judgement (Tschandl et al., 2020).

Let us briefly summarize some of the unexpected findings we encountered with:

- **Moderated Transparency**: Adding more AI transparency does not always help, can even worsen performance by causing higher cognitive load, complicating understanding and lowering decision speed.
- **Trust vs. Reliance**: Human trust does not equal human reliance on using AI. These two are not as closely correlated as one might expect.
- **Expertise as a Double-edged Sword**: While AI often supports less experienced users, its effectiveness is not universal, as expertise and confidence can both enhance and limit collaboration.
- **AI Performance vs Hybrid Performance:** Strong AI performance may not be the best contributor to the hybrid performance, on one hand lower quality AI can improve human engagement and performance, on the other complementarity may be of higher importance than accuracy.
- **Selective use of AI**: Not so perfect AI tools can still be useful if the human users are able to pick those instances where the AI appropriately contributes.

Overall, these findings present the complex, often non-linear interactions between key performance factors in human-AI collaboration. To achieve a successful collaboration requires more than just designing high-performing AI systems, it involves creating systems that can dynamically interact with human users in ways that enhance overall decision-making and teamwork context-specifically. Such systems embody the principles of HCAI by prioritizing user-centered design, trustworthiness, and ethical considerations, ensuring they enhance human capabilities in diverse, real-world scenarios.

6.3. Summary of contributions

This paper contributes to a better understanding of human-AI collaboration by looking closely at empirical evidence of factors that influence hybrid performance. We argue that understanding hybrid performance is challenging because AI, as a new form of automation, has unique features and the ability to collaborate directly with human users.

Several performance factors have been studied before, but there is no such performance framework grounded in empirical evidence, which addresses the complex problem space of hybrid performance comprehensively.

To address this gap, we conducted a qualitative meta-analysis of empirical AI implementation studies. We screened nearly 2000 papers (2018–2024) and fully analyzed 206 empirical studies focusing on performance in human-AI collaboration across various fields. This makes our paper and our findings timely and relevant for current AI developments.

Our key contributions are as follows:

Performance Factors: By completing a comprehensive meta-analysis, we identified 24 critical performance factors influencing hybrid performance, categorizing them into four groups: AI Factors, Human Factors, Interaction Factors, and Task-Related Factors. We bring together the 24 factors into one structured framework and we showed how these factors influence each other. We describe each factor to help better understand its role and its interrelationships with other factors. These factors offer an initial roadmap towards a structured, hybrid performance framework of human-AI collaboration.

Dynamic Interdependencies: Our work examines 326 directional links between these factors based on qualitative pairwise coding within the dataset. We uncover the complex, non-linear interdependencies between the factors. We create a dependency graph to visualize and illustrate these interrelationships. Within that we also highlight that some of these relationships may be non-linear, that sometimes leads to unexpected effects. We suggest that applying a dependency graph for mapping the factor relationships can offer both conceptual clarity and serve as a useful, visual tool during AI system design.

Context-Specific Design and Ongoing Design Support: Our findings show that AI systems must be adapted to the specific situation. It’s not enough to build technically strong systems, the AI system design must also reflect specific user needs, task complexity, and how key factors interact in that setting. Here, the continuous follow-up of these factors during AI implementation is also essential. Monitoring their changes over time helps improve system performance and ensures better alignment with real-world use.

Advancing HCAI Principles: Our study offers practical insights to operationalize HCAI principles, particularly in addressing challenges regarding the two critical factors: transparency and trust. This can help align AI design with human values and support effective collaboration.

7. Implications for AI design and collaboration strategies

Our findings have significant implications for both AI design and human-AI interaction strategies. To achieve effective collaboration, AI systems must go beyond high technical performance and engage with human users in ways that enhance decision-making, adaptability, and teamwork.

AI systems should accommodate varying user expertise and respond dynamically to changing scenarios. This adaptability is crucial to ensure that the systems remain effective and aligned with user needs across diverse applications.

Designers should consider and balance transparency with cognitive load, as too much information can overwhelm users. Tailoring transparency to user expertise and task complexity is critical for fostering trust and understanding without compromising usability.

Recognizing the dynamic relationships among performance factors can help avoid suboptimal outcomes and support more effective collaboration strategies.

8. Conclusions: future directions and final remarks

Future research should explore the nuanced relationships between performance factors further and test their implications in diverse real-

world scenarios. The focus should be on developing methodologies and tools that support the design of AI systems capable of dynamically adapting to user needs and contextual constraints. By building on the foundations established in our study, researchers and practitioners can create AI systems that truly enhance human potential and enable more effective collaboration across various domains.

This study lays the initial groundwork for advancing hybrid performance management and aligns with HCAI principles by offering a structured approach to understanding and designing AI systems for collaborative excellence. By bridging empirical evidence with practical applications, we provide actionable insights that can guide the design, implementation, and operation of AI systems that work seamlessly alongside human partners. Through this contribution, we hope to inspire further research and practice in the rapidly evolving field of human-AI collaboration.

CRediT authorship contribution statement

Dóra Göndöcs: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Szabolcs Horváth:** Writing – original draft, Visualization, Validation, Formal analysis. **Viktor Dörfler:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendices

Summary of Common Limitations in Empirical Studies

This table summarizes common limitations found in a set of 20 empirical studies randomly selected from our final dataset of 206 papers. The goal was to reflect critically on the strength of the evidence and highlight recurring methodological challenges in human–AI collaboration research.

Limitation	Summary
Limited Generalizability	Many studies focus on specific domains (e.g., pathology, law, customer service) or unique tools (e.g., the Merlin app), limiting how well findings transfer across contexts. These context-bound insights are useful for depth but challenge broad application or policy formation.
Small or Unbalanced Samples	A number of studies involved small participant pools, lacked diversity, or had skewed distributions (e.g., few experts or overrepresentation of students or MTurk workers), undermining the robustness and statistical power of their conclusions.
Ecological Validity Issues	Experimental tasks often did not reflect real-world conditions, relying on simulations, isolated interactions, or simplified environments that miss the complexity and unpredictability of actual human-AI workflows.
Lack of Longitudinal Data	Most research captured a single point in time or short-term interaction, offering limited insight into how human-AI collaboration, trust, and performance evolve over longer periods or repeated use.
Incomplete Human-AI Interaction	In many cases, the AI systems provided limited or static feedback, and users had minimal means to influence or adapt the AI, reducing the richness and potential of true bidirectional collaboration.
Anchoring or Bias Effects	Presenting AI suggestions early in decision-making tasks often introduced anchoring effects, making it difficult to distinguish between genuine agreement and influence-driven decisions, complicating interpretation of results.
Use of Proxy or Simplified Tasks	Studies frequently relied on proxy tasks or artificially simplified problems, which, while manageable for experiments, do not always map to the complexity, ambiguity, or stakes found in real professional settings.
AI Literacy and Onboarding Gaps	Several studies found users lacked training, context, or understanding of how to interpret AI outputs effectively, pointing to a need for better onboarding and user education to enable productive collaboration.
AI Transparency Limitations	Many systems used black-box models or lacked explainability features, reducing user trust, hindering diagnosis of errors, and preventing users from learning the model's reasoning or limitations.

The list of randomly selected studies analyzing limitations:

1. "Help Me Help the AI": Understanding How Explainability Can Support Human-AI Interaction (Kim et al., 2023a)
2. "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making (Cai et al., 2019)
3. A Human-AI Collaborative Approach for Clinical Decision Making on Rehabilitation Assessment (Lee et al., 2021)
4. A Missing Piece in the Puzzle: Considering the Role of Task Complexity in Human-AI Decision Making (Salimzadeh et al., 2023)
5. Adaptive Trust Calibration for Human-AI Collaboration (Kazuo Okamura and Seiji Yamada, 2020)
6. Adopting AI in the Context of Knowledge Work: Empirical Insights from German Organizations (von Richthofen et al., 2022)
7. Advancing Human-AI Complementarity: The Impact of User Expertise and Algorithmic Tuning on Joint Decision Making (Inkpen et al., 2023)
8. AI and AI-human based Salesforce Hiring using Conversational Interview Videos (Chakraborty et al., 2025)
9. AI Assistance in Legal Analysis: An Empirical Study (Choi and Schwarcz, 2025)
10. Algorithm-Augmented Work and Domain Experience: The Countervailing Forces of Ability and Aversion (Allen and Choudhury, 2021)
11. Artificial Intelligence Assistance Significantly Improves Gleason Grading of Prostate Biopsies by Pathologists (Bulten et al., 2021)
12. Artificial Intelligence versus Software Engineers (Nascimento et al., 2023)
13. Augmenting Medical Diagnosis Decisions? An Investigation into Physicians' Decision Making Process with Artificial Intelligence (Jussupow et al., 2021)
14. Batman and Robin in Healthcare Knowledge Work: Human-AI Collaboration by Clinical Documentation Integrity Specialists (Bossen and Pine, 2023)
15. Capturing Humans' Mental Models of AI: An Item Response Theory Approach (Kelly et al., 2023)
16. Cobots in Knowledge Work: Human-AI Collaboration in Managerial Professions (Sowa et al., 2021)
17. Cognitive Challenges in Human-Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation (Fügener et al., 2021a)

18. Collaborating with a Text-Based Chatbot - An Exploration of Real-World Collaboration Strategies Enacted during Human-Chatbot Interactions (Rapp et al., 2023)
19. Does the Whole Exceed its Parts? The Effect of AI Explanations on Complementary Team Performance (Bansal et al., 2021b)
20. Examining the Impact of Varying Levels of AI Teammate Influence on Human-AI Teams (Flathmann et al., 2023a)

Summary of Selected Empirical Cases

We find these real-world cases especially informative to our study since the empirical results of these papers typically cover a wider range of factors and observations compared to those focusing on a few, selected variables under controlled conditions. We have identified 10 papers that we consider highly relevant and we highlight their findings. It is important to note that our overall findings support the dynamic, complex, and often non-linear relationships between the factors, as outlined in the previous sections. These characteristics are clearly represented in the following specific studies as well.

How does artificial intelligence (AI) enhance hospitality employee innovation? The roles of exploration, AI trust, and proactive personality (Kong et al., 2024**)**

The study examines how the employee's innovation performance is influenced by the autonomy they perceive when using AI tools for their work. They use a survey-based method and collect data from 407 employees working in the hospitality sector. They find that "when employees hold higher AI trust and proactive personality, perceived AI-supported autonomy has a stronger relationship with innovative performance via exploration".

Human-Artificial Intelligence Collaboration in Prediction: A Field Experiment in the Retail Industry (Revilla et al., 2023)

This field experiment was carried out on data provided by an international FMCG company's 1888 units involving AI-driven demand forecasting. The paper found that the (predictive) context may require changes in the configuration of the human-AI interaction. "It shows that human judgment is likely to complement AIML-driven models when task environments are associated with long time horizons and low uncertainty. In contrast, our research suggests that this is not the case when the task requires coping with short time horizons and high uncertainty, in which case human judgment is unlikely to improve AIML-driven models." The results contribute to our findings that several factors are at play by "…demonstrating that the key issues relate not only to the configuration of the algorithm but also to how it is adaptively applied and integrated with human expertise in different contexts to reach the best performance".

Friend or Foe? Teaming Between Artificial Intelligence and Workers with Variation in Experience (W. Wang et al., 2023)

This paper is based on an AI tool used by medical coding experts in a publicly traded US company. In their work the researchers focus on the dynamics of the expertise in the context of human-AI interaction. Analyzing a large number of collaborative coding data they "show that experience is multidimensional and different dimensions have distinct interactions with AI and play unique roles in human-AI teaming." They suggest that "different types of experience have opposite effects on human-AI collaboration. Our task-based experience reflects the concept of ability, and our seniority-based experience is related to the aversion effect." Their results provide evidence that "greater task experience does not necessarily lead to algorithm aversion, but greater seniority experience could hamper the effective use of AI." Here is how: "Highly task-experienced coders enjoy greater improvements in productivity. However, when we turn to a different experience measure using seniority, we surprisingly find that junior coders experience greater gains in teaming with AI than senior coders. We attribute this to senior workers' greater sensitivity to AI imperfections, which results in their resistance to AI use and therefore lower benefits for them."

Humans, AI, and Context: Understanding End-Users' Trust in a Real-World Computer Vision Application (Kim et al., 2023b)

In this case study based on interviews with users of a popular, AI-based bird identification app, they delve into the interactions of trust, expertise, and AI-supported decision-making. They found that end-users' trust relationship with AI is complex. Instance-specific decisions about AI reliance "illustrate the rich reasoning behind participants' trust-related behaviors, where factors of trust interact with each other. Participants used their domain knowledge (human-related factor) to assess task difficulty (context-related factor) and weighted it against the app's ability (AI-related factor) to decide whether or not to accept the app output or adopt the app in the given situation." Expertise plays an important role since "Participants with domain knowledge assessed the correctness of an app output by, for example, cross-checking it with their own… Participants without domain knowledge, however, had difficulties in assessing the correctness of app outputs and, consequently, the app's overall ability."

Ignore, Trust, or Negotiate: Understanding Clinician Acceptance of AI-Based Treatment Recommendations in Health Care (Sivaraman et al., 2023)

This mixed-method study uses a novel AI-supported interface for sepsis treatment recommendations where intensive care clinicians made AI-assisted decisions on real patient cases. They find that users interacting with the AI system often move beyond the expectation "…that clinicians would calibrate their own certainty against the AI and make a binary decision about whether to accept its advice in each case. However, only a few participants (the Consider group, Sec. 5.2.3) exhibited this dichotomous form of reliance. Instead, most participants engaged in a more nuanced form of partial reliance on the AI, often involving a negotiation between the initial clinical assessment and various aspects of the recommendation." Therefore the explanations provided were generally found useful and building confidence still "explainable AI has complex effects on cognitive effort, especially when clinicians must decompose every recommendation into aspects with differing levels of credibility." The type and presentation of the explanation matters in these scenarios: "while the Feature Explanation chart helped participants decide how much weight to place on the recommendation overall, it did not support their ability to assign value to individual recommendation components. On the other hand, the Alternative Treatments approach may have better supported negotiation behaviors by allowing the AI to "present its findings" (P4) across a range of options." They also highlight that for those not integrating AI advice into their decision-making the AI "only served to lower their confidence in their decision-making and increase the perceived difficulty of the case". For those who do, "results might suggest that the visibility and complexity of the AI recommendations be adjusted based on the users' confidence or the discordance between their decision and the AI."

Exploring the Synergies in Human-AI Hybrids: A Longitudinal Analysis in Sales Forecasting (Fahse and Schmitt, 2023)

The research studied the context of an AI-supported next-day sales forecasting function in a bakery company running the system for more than 3 years already. As a general finding they confirm "previous research in the evaluation that humans tend to deteriorate AI forecasts when overriding". They also show that human interaction evolves over time finding that the "course of overriding changed over time in a U-shape." reaching a plateau. So human intervention might be related to how well the AI system is adapted to the specific (local) context: "…optimal forecasts in terms of certain performance metrics do not inevitably yield optimal real-life business outcomes. Complex human behavior, such as feeling uncomfortable when shelves are sporadically filled, could theoretically be incorporated into AI systems, but are not usually considered from the start… by analyzing the waste ratio and resulting empty shelves when waste ratios are too low, the results show that critical human thinking is still important in hybrid systems and should find its way into the system. AI systems are great for specific tasks, but do not always encompass all relevant features."

Algorithm-Augmented Work and Domain Experience: The Countervailing Forces of Ability and Aversion (Allen and Choudhury, 2021)

This study tests its hypothesis in a large Indian tech company's IT department resolving help-desk tickets with the use of algorithmic assistance. Similarly to "Friend or Foe? Teaming Between Artificial Intelligence and Workers with Variation in Experience" they examine countervailing forces of human experience arguing that "domain experience can increase performance via the ability to assess the quality of algorithmic advice (e.g., identify inaccurate predictions), but aversion may decrease performance via rejecting accurate algorithmic advice". They report that "low experience workers' low performance is driven by lack of ability to assess algorithmic advice, and high experience workers' failure to improve performance using algorithms is driven by their relatively high algorithmic aversion." At the same time "only moderately experienced workers performed significantly better when using the algorithm". They conclude that "whether an expert will do better or worse with an algorithm is not merely how much domain experience they have, but rather whether the aversion effect overpowers the ability effect in a given context".

Batman and Robin in Healthcare Knowledge Work: Human-AI Collaboration by Clinical Documentation Integrity Specialists (Bossen and Pine, 2023)

Similarly to the "Friend or Foe? Teaming Between Artificial Intelligence and Workers with Variation in Experience" research, the paper deals with the real-world collaboration between medical documentation specialists (CDIS) and an AI tool to support their job. The research included observations and interviews with 23 CDIS employees of the company operating six hospitals in the US. The CDIS find their collaboration with AI "a 'Batman and Robin' relationship: they are in charge, having the discretion of what and when to code, as well as when to query physicians in order to improve the documentation". They consider the probabilistic outputs of the AI software as "they have worked out which level of trust to grant the CAC [Computer Assisted Coding], as they know that the CAC does not catch all possible codes and suggests irrelevant codes. They have what Bansal et al. (2019a) call a mental model of the CAC's decision making, and are quite aware that an 'error boundary' exists, though with the CAC both are somewhat blurred and inconsistent." This paper emphasizes the importance of how the collaboration with AI will impact an existing workflow: "the AI in this case is embedded in a workflow and software that already is found to support and make CDIS' work easier and more efficient. Contrary to many instances in which AI is implemented as a new system into an existing workflow disrupting or slowing that workflow." They conclude that "this human-AI collaboration is contingent on several factors: the flexible integration of the AI into the workflow similar to the notion of unremarkable AI; supporting the CDIS' sensemaking; the CDIS' knowledge about the CAC being predictably unreliable; an experience by the CDIS of the AI's value; humans remaining in control; and ability to experiment with the AI, which spurs reflection and learning for these knowledge workers."

To Engage or Not to Engage with AI for Critical Judgments – How Professionals Deal with Opacity When Using AI for Medical Diagnosis (Lebovitz et al., 2022)

This in-depth field study covers three departments of a major US hospital and illustrates the "hindering effects of opacity" physicians are faced with when using AI tools in their diagnostic process. "In all three departments, professionals first formed initial knowledge claims and then considered the AI knowledge claim, which frequently conflicted with their initial claim. In all three departments, professionals experienced opacity of AI-in-use because they had no insight into the underlying reasoning of a given AI result, which in turn, heightened their experience of uncertainty." The cognitive burden of this uncertainty is compounded with the fact that "the AI tool provides opinions on every case, regardless of the professional's degree of uncertainty" contrary to how they solicit opinion from their colleagues ("on a regular but infrequent basis"). Only in one department physicians integrated AI successfully in their work by applying what they call "AI interrogation practices". The interrogation practices often require significant time investment from radiologists by carefully considering AI outputs: closely examining AI-flagged areas, gathering more context, going back-and-forth between their findings and the AI's, reviewing prior images, etc. "Enacting AI interrogation practices enabled professionals to reconcile the two knowledge claims (by overruling the AI claim, reflectively agreeing with it, or synthesizing the claims synergistically)". They provide insight into AI transparency: "Based on our study's findings, explainable or interpretable AI may enable but not guarantee that professionals are able to integrate AI knowledge (i.e., engaged augmentation)." Another insight was made regarding the time-accuracy trade-off: "Within engaged augmentation, experts invested additional time to reconcile the AI knowledge claims by enacting AI interrogation practices. This additional time may be justified by improvements to care quality, but it is unclear whether healthcare systems are willing to commit that additional time."

"Brilliant AI Doctor" in Rural China – Tensions and Challenges in AI-Powered CDSS Deployment (Wang et al., 2021)

The results of this study build on observations and interviews with physicians using an AI-supported clinical decision-making system in primary health care institutions of rural China. One of the key observations they make relevant to our study is that the AI software is not designed according to and not adapting to the actual physician's workflow. "The participants believe that 'Brilliant Doctor' is designed to support the 'textbook' diagnosis workflow, and they acknowledge such workflow is more rigorous and legitimate… However, it is simply impossible for these clinicians to follow the textbook workflow. To workaround the high patient volume and the constant interruptions in these clinics, clinicians had to multitask… Even when they were using the system, they used the 'Type-In' interaction approach rather than the 'Click-Through' approach, since the latter approach requires much longer time to complete a diagnosis process, even though ironically the 'Click-Through' approach was designed to make clinician's work easier and more efficient… Participants reported why the 'Click-Through' approach was rarely used. Not only that clicking through different pages and answering all the questions is time-consuming, but also the system forces the clinician to follow its clinical reasoning process, eroding the natural flow of clinicians' work routines and diagnostic practice."

Data availability

Data will be made available on request.

References

Abdel-Karim, B.M., Pfeuffer, N., Carl, K.V., Hinz, O., 2023. How AI-based systems can induce reflections: the case of AI-augmented diagnostic work. MIS Q. (4).

Acemoglu, D., Restrepo, P., 2018. Artificial intelligence, automation, and work. The Economics of Artificial Intelligence: An Agenda. University of Chicago Press, pp. 197–236. https://doi.org/10.7208/chicago/9780226613475.003.0008.

Acemoglu, D., Restrepo, P., 2019. Artificial intelligence, automation, and work. In: Agrawal, A., Gans, J., Goldfarb, A. (Eds.), The Economics of Artificial Intelligence: An Agenda. University of Chicago Press, pp. 197–236. http://www.nber.org/chapters/c14027.

Adam, M., Diebel, C., Goutier, M., Benlian, A., 2024. Navigating autonomy and control in Human-AI delegation: user responses to technology- versus user-invoked task allocation. Decis Support Syst 180, 114193. https://doi.org/10.1016/j.dss.2024.114193.

Aguinis, H., Pierce, C.A., Bosco, F.A., Dalton, D.R., Dalton, C.M., 2011. Debunking myths and urban legends about meta-analysis. Organ Res Methods 14 (2), 306–331. https://doi.org/10.1177/1094428110375720.

Allen, R., Choudhury, P., 2021. Algorithm-augmented work and domain experience: the countervailing forces of ability and aversion. Organ. Sci. https://doi.org/10.1287/orsc.2021.1554.

Altintas, O., Seidmann, A., & Gu, B. (2023). The effect of interpretable artificial intelligence on repeated managerial decision-making under uncertainty. *Available at SSRN 4331145*. https://doi.org/10.2139/ssrn.4331145.

Amershi, S., Cakmak, M., Knox, W.B., Kulesza, T., 2014. Power to the people: the role of humans in interactive machine learning. AI Mag. 35 (4), 105–120. https://doi.org/10.1609/aimag.v35i4.2513.

Andre, F., Fortner, P., Aurich, M., Seitz, S., Jatsch, A.-K., Schöbinger, M., Wels, M., Kraus, M., Gülsün, M.A., Frey, N., Sommer, A., Görich, J., Buss, S.J., 2023. Human AI teaming for coronary ct angiography assessment: impact on imaging workflow and diagnostic accuracy. Diagnostics 13 (23), 3574. https://www.mdpi.com/2075-4418/13/23/3574.

Ardichvili, A., Dirani, K., Jabarkhail, S., El Mansour, W., Aboulhosn, S., 2024. Using generative AI in Human resource development: an applied research study. Hum. Resour. Dev. Int. 27 (3), 388–409. https://doi.org/10.1080/13678868.2024.2337964.

Bainbridge, L., 1983. Ironies of automation. In: Johannsen, G., Rijnsdorp, J.E. (Eds.), Analysis, Design and Evaluation of Man–Machine Systems. Pergamon Press, pp. 129–135. https://doi.org/10.1016/B978-0-08-029348-6.50026-9.

Bansal, G., Nushi, B., Kamar, E., Lasecki, W., Weld, D., Horvitz, E., 2019a. Beyond accuracy: the role of mental models in Human-AI team performance. In: Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, 7, pp. 2–11. https://doi.org/10.1609/hcomp.v7i1.5285.

Bansal, G., Nushi, B., Kamar, E., Weld, D.S., Lasecki, W.S., Horvitz, E., 2019b. Updates in Human-AI teams: understanding and addressing the performance/compatibility tradeoff. In: AAAI Conference on Artificial Intelligence.

Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M.T., & Weld, D. (2021). Does the whole exceed its parts? The effect of AI explanations on complementary team performance.

Bansal, G., Nushi, B., Kamar, E., Horvitz, E., Weld, D.S., 2021a. Is the most accurate AI the best teammate? Optimizing AI for teamwork. In: AAAI Conference on Artificial Intelligence.

Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M.T., & Weld, D. (2021b, 2021). Does the whole exceed its parts? The effect of AI explanations on complementary team performance.

Bartels, E.M., 2013. How to perform a systematic search. Best Pract. Res. Clin. Rheumatol. 27 (2), 295–306.

Baryannis, G., Dani, S., Antoniou, G., 2019. Predicting supply chain risks using machine learning: the trade-off between performance and interpretability. Future Gener. Comput. Syst. 101, 993–1004. https://doi.org/10.1016/j.future.2019.07.059.

Bas, A., Sinclair, M., Dörfler, V., 2023. Sensing: the elephant in the room of management learning. Manag. Learn. 54 (4), 489–510. https://doi.org/10.1177/13505076221077226.

Beese, J., Fahse, T., 2023, 11-16 June 2023. In: *Modern Centaurs: How Humans and AI Systems Interact in Sales Forecasting* ECIS 2023: 31st European Conference on Information Systems. Norway. Kristiansand. https://aisel.aisnet.org/ecis2023_rp/267.

Berberian, B., Le Guillou, M., Pagliari, M., 2023. Communicating AI intentions to boost Human AI cooperation. In: Workshop" MULTITTRUST-Multidisciplinary Perspectives on Human-AI Team Trust" at the 2nd International Conference on Hybrid Human-Artificial Intelligence (HHAI-WS 2023).

Bertrand, A., Viard, T., Belloum, R., Eagan, J.R., Maxwell, W.. *On selective, mutable and dialogic Xai: a review of what users say about different types of interactive explanations*. https://doi.org/10.1145/3544548.3581314.

Bhattacharya, A., Stumpf, S., & Verbert, K. (2024, 2024). An explanatory model steering system for collaboration between domain experts and AI.

Bienefeld, N., Kolbe, M., Camen, G., Huser, D., Buehler, P.K., 2023. Human-AI teaming: leveraging transactive memory and speaking up for enhanced team effectiveness. Front Psychol 14, 1208019. https://doi.org/10.3389/fpsyg.2023.1208019.

Billings, C.E. (1997). Aviation automation the search for a Human centered approach Lawrence Erlbaum Associates. *Inc., Mahwah, NJ*. https://doi.org/10.1201/9781315137995.

Blaurock, M., Büttgen, M., Schepers, J., 2024. Designing collaborative intelligence systems for employee-AI service Co-production. J. Serv. Res. 0 (0), 10946705241238751. https://doi.org/10.1177/10946705241238751.

Bondi, E., Koster, R., Sheahan, H., Chadwick, M., Bachrach, Y., Cemgil, T., Paquet, U., Dvijotham, K., 2022. Role of Human-AI interaction in selective prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence.

Bornet, P., Wirtz, J., Davenport, T.H., De Cremer, D., Evergreen, B., Fersht, P., Gohel, R., Khiyara, S., Mullakara, N., Sund, P., 2025. Agentic Artificial Intelligence: Harnessing AI Agents to Reinvent Business, Work and Life. Irreplaceable Publishing. https://books.google.co.uk/books?id=UfRTEQAAQBAJ.

Bossen, C., Pine, K.H., 2023. Batman and Robin in healthcare knowledge work: human-AI collaboration by clinical documentation integrity specialists. ACM Trans. Comput.-Hum. Interact. 30 (2), 1–29. https://doi.org/10.1145/3569892.

Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qual Res Psychol 3 (2), 77–101. https://doi.org/10.1191/1478088706qp063oa.

Braun, V., Clarke, V., 2012. Thematic analysis. Apa Handbook of Research Methods in Psychology, Vol 2: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological. American Psychological Association, pp. 57–71. https://doi.org/10.1037/13620-004.

Braun, V., Clarke, V., 2014. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 9, 1–2. https://doi.org/10.3402/qhw.v9.26152.

Braun, V., Clarke, V., 2019. Reflecting on reflexive thematic analysis. Qual. Res. Sport Exerc. Health 11 (4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806.

Braun, M., Greve, M., Riquel, J., Brendel, A., Kolbe, L., 2022, 18-24 June 2022. In: *Meet Your New Colle(AI)Gue – Exploring the Impact of Human-AI Interaction Designs on User Performance* ECIS 2022: 30th European Conference on Information Systems. Timişoara, Romania. https://aisel.aisnet.org/ecis2022_rp/122.

Braun, M., Greve, M., Brendel, A.B., Kolbe, L.M., 2023. Humans supervising artificial intelligence – Investigation of designs to optimize error detection. J. Decis. Syst. 1–26. https://doi.org/10.1080/12460125.2023.2260518.

Brehmer, M., Leibig, C., Bunk, S., Byng, D., Pinker, K., Umutlu, L., 2022. Articles combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. Lancet 4. https://doi.org/10.1016/S2589-7500(22)00070-X.

Brickson, S., 2021. Toward the humanistic manager and transformative Management & Organizing: editors panel & let's go! (Session 160). https://2021.aom.org/meetings/virtual/dcGsdwFNtJj2piQwi.

Brynjolfsson, E., McAfee, A., 2014. The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. W. W. Norton & Company. https://books.google.co.uk/books?id=9sMeBwAAQBAJ.

Buçinca, Z., Malaya, M.B., Gajos, K.Z., 2021. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. In: Proc. ACM Hum.-Comput. Interact., 5. https://doi.org/10.1145/3449287. Article 188.

Buchanan, B.G., Sutherland, G.L., Feigenbaum, E.A., 1969. Heuristic dendral: a program for generating explanatory hypotheses in organic chemistry. In: Meltzer, B., Michie, D. (Eds.), Machine Intelligence. Oliver & Boyd, pp. 209–254.

Bui, L., Pezzola, M., Bandara, D., 2023. In: Degen, H., Ntoa, S. (Eds.), How Do AI Explanations Affect Human-AI Trust?. Artificial Intelligence in HCI, Cham.

Bulten, W., Balkenhol, M., Belinga, J.-J.A., Brilhante, A., Çakır, A., Egevad, L., Eklund, M., Farré, X., Geronatsiou, K., Molinié, V., 2021. Artificial intelligence assistance significantly improves gleason grading of prostate biopsies by pathologists. Mod. Pathol. 34 (3), 660–671. https://doi.org/10.1038/s41379-020-0640-y.

Bussone, A., Stumpf, S., O'Sullivan, D., 2015. The role of explanations on trust and reliance in clinical decision support systems. In: 2015 International Conference on Healthcare Informatics.

Cabitza, F., Campagner, A., Sconfienza, L.M., 2021. Studying Human-AI collaboration protocols: the case of the Kasparov's law in radiological double reading. Health Inf Sci Syst 9 (1), 8. https://doi.org/10.1007/s13755-021-00138-8.

Cabitza, F., Campagner, A., Ronzio, L., Cameli, M., Mandoli, G.E., Pastore, M.C., Sconfienza, L.M., Folgado, D., Barandas, M., Gamboa, H., 2023. Rams, hounds and white boxes: investigating Human–AI collaboration protocols in medical diagnosis. Artif Intell Med 138, 102506. https://doi.org/10.1016/j.artmed.2023.102506.

Cabitza, F., Fregosi, C., Campagner, A., Natali, C., 2024. In: Longo, L., Lapuschkin, S., Seifert, C. (Eds.), Explanations Considered Harmful: The Impact Of misleading Explanations On Accuracy In hybrid Human-AI Decision Making. Explainable Artificial Intelligence, Cham.

Cabrera, A., Perer, A., Hong, J., 2023. Improving Human-AI collaboration with descriptions of AI behavior. Proc. ACM Hum.-Comput. Interact. 7, 1–21. https://doi.org/10.1145/3579612.

Cai, C.J., Winter, S., Steiner, D., Wilcox, L., Terry, M., 2019. Hello AI": uncovering the onboarding needs of medical practitioners for Human-AI collaborative decision-making. In: Proc. ACM Hum.-Comput. Interact, 3. https://doi.org/10.1145/3359206. Article 104.

Calisto, F.M., Santiago, C., Nunes, N., Nascimento, J.C., 2021. Introduction of Human-centric AI assistant to aid radiologists for multimodal breast image classification. Int J Hum Comput Stud 150, 102607. https://doi.org/10.1016/j.ijhcs.2021.102607.

Cao, S., Gomez, C., Huang, C.-M., 2023. How time pressure in different phases of decision-making influences Human-AI collaboration. Proc. ACM Hum.-Comput. Interact. 7 (CSCW2), 1–26. https://doi.org/10.1145/3610068.

Cao, S., Liu, A., Huang, C.-M., 2024. Designing for appropriate reliance: the roles of AI uncertainty presentation, initial user decision, and user demographics in AI-assisted decision-making. Proc. ACM Hum.-Comput. Interact. 8 (CSCW1), 1–32. https://doi.org/10.1145/3637318.

Cassenti, D.N., Kaplan, L.M., Roy, A., 2023. Representing uncertainty information from AI for Human understanding. In: , 67, pp. 177–182. https://doi.org/10.1177/21695067231193649.

Cecil, J., Lermer, E., Hudecek, M.F.C., Sauer, J., Gaube, S., 2024. Explainability does not mitigate the negative impact of incorrect AI advice in a personnel selection task. Sci Rep 14 (1), 9736. https://doi.org/10.1038/s41598-024-60220-5.

Chakraborty, I., Chiong, K., Dover, H., Sudhir, K., 2025. Can AI and AI-hybrids detect persuasion skills? Salesforce hiring with conversational video interviews. Mark. Sci. 44 (1), 30–53. https://doi.org/10.1287/mksc.2023.0149.

Chen, V., Liao, Q.V., Wortman Vaughan, J., Bansal, G., 2023. Understanding the role of Human intuition on reliance in Human-AI decision-making with explanations. In: , 7, pp. 1–32. https://doi.org/10.1145/3610219.

Cheng, F., Zouhar, V., Arora, S., Sachan, M., Strobelt, H., El-Assady, M., 2024. Relic: investigating large language model responses using self-consistency. Proc. CHI Conf. Hum. Factors Comput. Syst.

Chiang, C.-W., Yin, M., 2021. In: *You'd Better Stop! Understanding Human Reliance on Machine Learning Models under Covariate Shift* WebSci '21: 13th ACM Web Science Conference, Virtual Event. United Kingdom. https://doi.org/10.1145/3447535.3462487.

Chiang, C.-W., & Yin, M. (2022, 2022). Exploring the effects of machine learning literacy interventions on laypeople's reliance on machine learning models.

Chiang, C.-W., Lu, Z., Li, Z., & Yin, M. (2024, 2024). Enhancing AI-assisted group decision making through llm-powered Devil's advocate.

Choi, J., Schwarcz, D., 2025. AI assistance in legal analysis: an empirical study. J Leg. Educ 73 (2), 384–420. https://doi.org/10.2139/ssrn.4539836.

Choi, E., Bahadori, M.T., Schuetz, A., Stewart, W.F., Sun, J., 2016. Doctor AI: predicting clinical events via recurrent neural networks. In: MLHC 2016: 1st Machine Learning and Healthcare Conference. Los Angeles, CA. https://doi.org/10.48550/arXiv.1511.05942.

Chong, L., Zhang, G., Goucher-Lambert, K., Kotovsky, K., Cagan, J., 2022. Human confidence in artificial intelligence and in themselves: the evolution and impact of confidence on adoption of AI advice. Comput Hum. Behav 127, 107018. https://doi.org/10.1016/j.chb.2021.107018.

Chong, L., Raina, A., Goucher-Lambert, K., Kotovsky, K., Cagan, J., 2023. The evolution and impact of Human confidence in artificial intelligence and in themselves on AI-assisted decision-making in design. J. mech. des. 145 (3), 031401. https://doi.org/10.1115/1.4055123.

Chong, L., Kotovsky, K., Cagan, J., 2024. Human designers' Dynamic confidence and decision-making when working with more than one artificial intelligence. J. Mech. Des. 146 (8). https://doi.org/10.1115/1.4064565.

Choy, G., Khalilzadeh, O., Michalski, M., Do, S., Samir, A.E., Pianykh, O.S., Geis, J.R., Pandharipande, P.V., Brink, J.A., Dreyer, K.J., 2018. Current applications and future impact of machine learning in radiology. Radiology 288 (2), 318.

Coeckelbergh, M., 2020. Should we treat Teddy Bear 2.0 as a kantian dog? Four arguments for the indirect moral standing of personal social robots, with implications for thinking about animals and humans. Minds Mach. https://doi.org/10.1007/s11023-020-09554-3.

Coombs, C., Hislop, D., Taneva, S.K., Barnard, S., 2020. The strategic impacts of intelligent automation for knowledge and service work: an interdisciplinary review. J. Strateg. Inf. Syst. 29 (4), 101600. https://doi.org/10.1016/j.jsis.2020.101600.

Corvelo Benz, N.L., Rodriguez, M.G., 2023. Human-aligned calibration for AI-assisted decision making. In: NIPS '23: 37th International Conference on Neural Information Processing Systems. New Orleans, LA. https://dl.acm.org/doi/10.5555/3666122.3666765.

Cvetkovic, I., & Bittner, E.A. (2022). Task delegability to AI: evaluation of a framework in a knowledge work context. HICSS.

Dörfler, V., Stierand, M., 2021. Bracketing: a phenomenological theory applied through transpersonal reflexivity. J. Organ. Change Manag. 34 (4), 778–793. https://doi.org/10.1108/JOCM-12-2019-0393.

Dörfler, V., 2022. What Every CEO Should Know About AI. Cambridge University Press. https://doi.org/10.1017/9781009037853.

Dörfler, V., 2023. *Critical Interpretivism: The First Outline* PHILOS 2023: 3rd Colloquium on Philosophy and Organization Studies, 22-24 June 2023. Chania, Greece.

Dang, H., Goller, S., Lehmann, F., Buschek, D., 2023. Choice over control: how users write with large language models using diegetic and *non-diegetic prompting*. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. Hamburg, Germany. https://doi.org/10.1145/3544548.3580969.

Danry, V., Pataranutaporn, P., Mao, Y., Maes, P., 2023. Don't just tell me, ask me: AI systems that intelligently frame explanations as questions improve Human logical discernment accuracy over causal AI explanations. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. Hamburg, Germany. https://doi.org/10.1145/3544548.3580672.

Davenport, T.H., Glover, W.J., 2018. Artificial intelligence and the augmentation of health care decision-making. NEJM Catal. 4 (3). https://catalyst.nejm.org/ai-technologies-augmentation-healthcare-decisions/.

Davenport, T.H., Harris, J.G., 2017. Competing On Analytics: The New Science of Winning, 2nd ed. Harvard Business School Press https://books.google.co.uk/books?id=EUDkAQAACAAJ.

Davenport, T.H., Kirby, J., 2015. Beyond automation. Harv Bus Rev 93 (6), 58–65.

Davenport, T.H., Kirby, J., 2016. Only Humans Need Apply: Winners and Losers in the Age of Smart Machines. Harper Business, New York, NY.

Davenport, T.H., Harris, J.G., Morison, R., 2010. Analytics At Work: Smarter Decisions, Better Results. Harvard Business Press. https://books.google.hu/books?id=2otJuvfvflgC.

Davenport, T.H., 2018. From analytics to artificial intelligence. J. Bus. Anal. 1 (2), 73–80.

Davenport, T.H., 2023. Generative AI for customer service at Ada and Wealthsimple. Forbes. https://www.forbes.com/sites/tomdavenport/2023/09/20/generative-ai-for-customer-service-at-ada-and-wealthsimple/?sh=146e5249178e.

Daws, R., 2020. Medical chatbot using Openai's gpt-3 told a fake patient to kill themselves. AI News. https://artificialintelligence-news.com/2020/10/28/medical-chatbot-openai-gpt3-patient-kill-themselves/.

De Brito Duarte, R., 2023. Towards responsible AI: developing explanations to increase Human-AI collaboration. Hhai 2023: Augmenting Human Intellect. IOS Press, pp. 470–482. https://doi.org/10.3233/FAIA230126.

Dearden, A., Harrison, M., Wright, P., 2000. Allocation of function: scenarios, context and the economics of effort. Int J Hum Comput Stud 52 (2), 289–318. https://doi.org/10.1006/ijhc.1999.0290.

Dekker, S.W.A., Woods, D.D., 2002. Maba-Maba or Abracadabra? Progress on Human–Automation Co-ordination. Cogn. Technol. Work 4 (4), 240–244. https://doi.org/10.1007/s101110200022.

Dell'Acqua, F., Kogut, B., Perkowski, P., 2025. Super Mario meets AI: experimental effects of automation and skills on team performance and coordination. Rev Econ Stat 107 (4), 951–966. https://doi.org/10.1162/rest_a_01328.

Dell'Acqua, F., 2022. *Falling asleep at the wheel: human/AI collaboration in a field experiment on hr recruiters* (Working Papers). https://anacanhoto.com/wp-content/uploads/2024/08/554ee-fallingasleepatthewheel-fabriziodellacqua.pdf.

Dell'Acqua, F., McFowland, E.I., Mollick, E.R., Lifshitz-Assaf, H., Kellogg, K.C., Rajendran, S., Krayer, L., Candelon, F., & 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, Issue 24–013). https://doi.org/10.2139/ssrn.4573321.

Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., & Ebel, P. (2021). The future of Human-AI collaboration: a taxonomy of design knowledge for Hybrid Intelligence systems. *arXiv preprint arXiv:2105.03354*. https://doi.org/10.48550/arXiv.2105.03354.

Dennett, D.C., 1998. When Hal kills, who's to blame? Computer ethics. In: Stork, D.G. (Ed.), Hal's Legacy: 2001′s Computer as Dream and Reality. MIT Press, pp. 351–365. https://doi.org/10.7551/mitpress/3404.003.0018.

Dhuliawala, S., Zouhar, V., El-Assady, M., & Sachan, M. (2023). A diachronic perspective on user trust in AI under uncertainty. *arXiv preprint arXiv:2310.13544*. https://doi.org/10.48550/arXiv.2310.13544.

Dwivedi, Y.K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P.V., Janssen, M., Jones, P., Kar, A.K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., Medaglia, R., Le Meunier-FitzHugh, K., Le Meunier-FitzHugh, L.C., Misra, S., Mogaji, E., Sharma, S.K., Singh, J.B., Raghavan, V., Raman, R., Rana, N.P., Samothrakis, S., Spencer, J., Tamilmani, K., Tubadji, A., Walton, P., Williams, M.D, 2021. Artificial Intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. Int J Inf Manage 57, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002.

Ebben, M., 2022. Automation and augmentation: human labor as essential complement to machines. In: Association, I.R.M. (Ed.), Research Anthology on Cross-Disciplinary Designs and Applications of Automation. IGI Global, pp. 1–19. https://doi.org/10.4018/978-1-6684-3694-3.ch001.

Ebermann, C., Selisky, M., Weibelzahl, S., 2023. Explainable AI: the effect of contradictory decisions and explanations on users' Acceptance of AI systems. Int. J. Hum.–Comput. Interact. 39 (9), 1807–1826. https://doi.org/10.1080/10447318.2022.2126812.

Eini-Porat, B., Amir, O., Eytan, D., Shalit, U., 2022. Tell me something interesting: clinical utility of machine learning prediction models in the Icu. J Biomed Inf. 132, 104107. https://doi.org/10.1016/j.jbi.2022.104107.

Emery, F.E., Trist, E.L., 1965. The causal texture of organizational environments. Hum. Relat. 18 (1), 21–32. https://doi.org/10.1177/001872676501800103.

Endsley, M.R., 2017. From here to autonomy: lessons learned from Human–Automation research. Hum Factors 59 (1), 5–27. https://doi.org/10.1177/0018720816681350.

Erlei, A., Sharma, A., & Gadiraju, U. (2024, 2024). Understanding choice independence and error types in Human-AI collaboration.

Fügener, A., Grahl, J., Gupta, A., Ketter, W., 2021a. Cognitive challenges in Human–Artificial intelligence collaboration: investigating the path toward productive delegation. Inf. Syst. Res. 33. https://doi.org/10.1287/isre.2021.1079.

Fügener, A., Grahl, J., Gupta, A., Ketter, W., 2021b. Will humans-in-the-loop become borgs? Merits and pitfalls of working with AI. Manag. Inf. Syst. Q. (MISQ)- Vol 45.

Fügener, A., Grahl, J., Gupta, A., Ketter, W., & Taudien, A. (2021c). Exploring user heterogeneity in Human delegation behavior towards AI. ICIS.

Fahse, T.B., & Schmitt, A. (2023). Exploring the synergies in Human-AI hybrids: a longitudinal analysis in sales forecasting. Exploring the Synergies in Human-AI Hybrids: A Longitudinal Analysis in Sales Forecasting.

Feigenbaum, E.A., 1977. The art of artificial intelligence: I. Themes and case studies of knowledge engineering. In: 5th International Joint Conference on Artificial Intelligence. Cambridge, MA. https://scholar.google.com/scholar_lookup?title=The%20art%20of%20artificial%20intelligence%3A%20I.%20Themes%20and%20case%20studies%20of%20knowledge%20engineering&publication_year=1977&author=E.%20Feigenbaum.

Feigenbaum, E.A., 1992. *A personal view of expert systems: looking back and looking ahead* (KSL 92-41).). https://purl.stanford.edu/gr891tb5766.

Fisher, G., Mayer, K., Morris, S., 2021. From the editors—Phenomenon-based theorizing. Acad. Manag. Rev. 46 (4), 631–639. https://doi.org/10.5465/amr.2021.0320.

Flathmann, C., Schelble, B., Rosopa, P., McNeese, N., Mallick, R., Chalil Madathil, K., 2023a. Examining the impact of varying levels of AI teammate influence on Human-AI teams. Int J Hum Comput Stud 177, 103061. https://doi.org/10.1016/j.ijhcs.2023.103061.

Flathmann, C., Schelble, B.G., McNeese, N.J., Knijnenburg, B., Gramopadhye, A.K., Chalil Madathil, K., 2023b. The purposeful presentation of AI teammates: impacts on Human acceptance and perception. Int. J. Hum.–Comput. Interact. 1–18. https://doi.org/10.1080/10447318.2023.2254984.

Floridi, L., Sanders, J.W., 2004. On the morality of artificial agents. Minds Mach. 14 (3), 349–379. https://doi.org/10.1023/B:MIND.0000035461.63578.9d.

Fogliato, R., Chappidi, S., Lungren, M., Fisher, P., Wilson, D., Fitzke, M., Parkinson, M., Horvitz, E., Inkpen, K., & Nushi, B. (2022). Who goes first? Influences of Human-AI workflow on decision making in clinical imaging.

Ford, S., Despeisse, M., 2016. Additive manufacturing and sustainability: an exploratory study of the advantages and challenges. J Clean Prod 137, 1573–1587. https://doi.org/10.1016/j.jclepro.2016.04.150.

Fukuchi, Y., Yamada, S., 2023. Dynamic selection of reliance calibration cues with AI reliance model. IEEE Access 11, 138870–138881. https://doi.org/10.1109/ACCESS.2023.3339548.

Göndöcs, D., Dörfler, V., 2024. AI in medical diagnosis: AI prediction & Human judgement. Artif Intell Med 149, 102769. https://doi.org/10.1016/j.artmed.2024.102769.

Gao, R., Saar-Tsechansky, M., De-Arteaga, M., Han, L., Lee, M.K., Lease, M., 2021. Human-AI collaboration with Bandit Feedback. Int. Jt. Conf. Artif. Intell.

Gentile, D., Donmez, B., Jamieson, G.A., 2023. Human performance consequences of normative and contrastive explanations: an experiment in machine learning for reliability maintenance. Artif Intell 321, 103945. https://doi.org/10.1016/j.artint.2023.103945.

Gilpin, L.H., Bau, D., Yuan, B.Z., Bajwa, A., Specter, M., & Kagal, L. (2018, 2018). Explaining explanations: an overview of interpretability of machine learning.

Glauben, A., Zöll, A., Sharma, B., Kristo, F., 2024, 1-5 July 2024. In: *The Dialog Trap: Exploring Potentially Detrimental Effects of Dialog-Based Interfaces for Generative AI*

Content Creation PACIS 2024: Pacific Asia Conference on Information Systems. Ho Chi Minh City, Vietnam. https://aisel.aisnet.org/pacis2024/track01_aibussoc/track01_aibussoc/15.

Glienke, M., Hartwich, N.J., Antons, D., 2023. Working with AI: how attitudes shape Human-AI collaboration. https://aisel.aisnet.org/icis2023/hti/hti/12.

Goldberg, K., 2012. What is automation? IEEE Trans. Autom. Sci. Eng. 9 (1), 1–2. https://doi.org/10.1109/TASE.2011.2178910.

Gomez, C., Cho, S.M., Huang, C.-M., & Unberath, M. (2023a). Designing AI support for Human involvement in AI-assisted decision making: a taxonomy of Human-AI interactions from a systematic review. *arXiv preprint arXiv:2310.19778*. https://doi.org/10.48550/arXiv.2310.19778.

Gomez, C., Unberath, M., Huang, C.-M., 2023b. Mitigating knowledge imbalance in AI-advised decision-making through collaborative user involvement. Int J Hum Comput Stud 172, 102977. https://doi.org/10.1016/j.ijhcs.2022.102977.

Gozalo-Brizuela, R., & Garrido-Merchan, E.C. (2023). Chatgpt is not all you need. A State of the art review of large generative AI models. *arXiv*, 1801.00631. https://doi.org/10.48550/arXiv.2301.04655.

Grisold, T., Schneider, J., 2023. Dynamics of Human-AI delegation in organizational routines. https://www.researchgate.net/publication/374914264_Dynamics_of_Human-AI_Delegation_in_Organizational_Routines?enrichId=rgreq-8e4938ee7aaa995bb1a1fa351436afde-XXX&enrichSource=Y292ZXJQYWdlOzM3NDkxNDI2NDtBUzoxMTQzMTI4MTIwMDY2NDA4N0AxNjk4MDUzOTE0MDQ5&el=1_x_2&_esc=publicationCoverPdf.

Gu, H., Huang, J., Hung, L., Chen, X., 2021. Lessons learned from designing an AI-enabled diagnosis tool for pathologists. Proc. ACM Hum.-Comput. Interact. 5, 1–25. https://doi.org/10.1145/3449084.

Gu, H., Yang, C., Haeri, M., Wang, J., Tang, S., Yan, W., He, S., Williams, C.K., Magaki, S., & Chen, X.A. (2023). Augmenting pathologists with Navipath: design and evaluation of a Human-AI collaborative navigation system.

Gu, K., Shang, R., Althoff, T., Wang, C., Drucker, S.M., 2024. How do analysts understand and verify AI-assisted data analyses?. In: Proceedings of the CHI Conference on Human Factors in Computing Systems.

Guggemos, J., 2024. To fear or not to fear – Human resource development professionals' Positioning towards artificial intelligence with a focus on augmentation. Comput. Educ.: Artif. Intell. 7, 100260. https://doi.org/10.1016/j.caeai.2024.100260.

Guo, E., Hao, K., 2020. This is the Stanford vaccine algorithm that left out frontline doctors. MIT Technol. Rev. https://www.technologyreview.com/2020/12/21/1015303/.

Guo, J., Mohanty, V., Ono, J.H.P., Hao, H., Gou, L., Ren, L., 2024. Investigating interaction modes and user agency in Human-llm collaboration for domain-specific data analysis. In: Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3613905.3651042.

Guzdial, M., Liao, N., Chen, J., Chen, S.-Y., Shah, S., Shah, V., Reno, J., Smith, G., Riedl, M.O., 2019. Friend, collaborator, student, manager: how design of an AI-driven game level editor affects creators. In: CHI '19: 2019 CHI Conference on Human Factors in Computing Systems. Glasgow, UK. https://doi.org/10.1145/3290605.3300854.

Hüllmann, J.A., Precht, H., & Wübbe, C. (2023). Configurations of Human-AI work in agriculture.

Haesevoets, T., De Cremer, D., Dierckx, K., Van Hiel, A., 2021. Human-machine collaboration in managerial decision making. Comput Hum. Behav 119, 106730. https://doi.org/10.1016/j.chb.2021.106730.

Harbarth, L., Gößwein, E., Bodemer, D., Schnaubert, L., 2024. (Over)Trusting AI recommendations: how system and person variables affect dimensions of complacency. Int. J. Hum.–Comput. Interact. 41 (1), 391–410. https://doi.org/10.1080/10447318.2023.2301250.

Hauptman, A.I., Schelble, B.G., Duan, W., Flathmann, C., McNeese, N.J., 2024. Understanding the influence of AI autonomy on AI explainability levels in Human-AI teams using a mixed methods approach. Cogn. Technol. Work 26 (3), 435–455. https://doi.org/10.1007/s10111-024-00765-7.

He, G., Kuiper, L., Gadiraju, U., 2023. Knowing about Knowing: an illusion of Human competence can hinder appropriate reliance on AI systems. In: CHI '23: 2023 CHI Conference on Human Factors in Computing Systems. Hamburg, Germany. https://doi.org/10.1145/3544548.3581025.

Heaven, W.D., 2019. Deep tbodenrouble for Deep learning. Nature 574 (7777), 163–166. https://doi.org/10.1038/d41586-019-03013-5.

Hekler, A., Utikal, J.S., Enk, A.H., Hauschild, A., Weichenthal, M., Maron, R.C., Berking, C., Haferkamp, S., Klode, J., Schadendorf, D., Schilling, B., Holland-Letz, T., Izar, B., von Kalle, C., Fröhling, S., Brinker, T.J., Schmitt, L., Peitsch, W.K., Hoffmann, F., Becker, J.C., Drusio, C., Jansen, P., Lodde, G., Sammet, S., Sondermann, W., Ugurel, S., Zader, J., Enk, A., Salzmann, M., Schäfer, S., Schäkel, K., Winkler, J., Wölbing, P., Asper, H., Bohne, A.-S., Brown, V., Burba, B., Deffaa, S., Dietrich, C., Dietrich, M., Drerup, K.A., Egberts, F., Erkens, A.-S., Greven, S., Harde, V., Jost, M., Kaeding, M., Kosova, K., Lischner, S., Maagk, M., Messinger, A.L., Metzner, M., Motamedi, R., Rosenthal, A.-C., Seidl, U., Stemmermann, J., Torz, K., Velez, J.G., Haiduk, J., Alter, M., Bär, C., Bergenthal, P., Gerlach, A., Holtorf, C., Karoglan, A., Kindermann, S., Kraas, L., Felcht, M., Gaiser, M.R., Klemke, C.-D., Kurzen, H., Leibing, T., Müller, V., Reinhard, R.R., Utikal, J., Winter, F., Eicher, L., Hartmann, D., Heppt, M., Kilian, K., Krammer, S., Lill, D., Niesert, A.-C., Oppel, E., Sattler, E., Senner, S., Wallmichrath, J., Wolff, H., Gesierich, A., Giner, T., Glutsch, V., Kerstan, A., Presser, D., Schrüfer, P., Schummer, P., Stolze, I., Weber, J., Drexler, K., Mickler, M., Stauner, C.T., Thiem, A., 2019. Superior skin cancer classification by the combination of Human and artificial intelligence. Eur J Cancer 120, 114–121. https://doi.org/10.1016/j.ejca.2019.07.019.

Hemmer, P., Westphal, M., Schemmer, M., Vetter, S., Vössing, M., & Satzger, G. (2023). *Human-AI collaboration: the effect of AI delegation on Human task performance and task satisfaction*. https://doi.org/10.48550/arXiv.2303.09224.

Henkenjohann, R., Trenz, M., 2024. Challenges in collaboration with generative AI: interaction patterns, outcome quality and perceived responsibility. In: ECIS 2024: 32nd European Conference on Information Systems. Paphos, Cyprus. https://aisel.aisnet.org/ecis2024/track05_fow/track05_fow/3.

Herm, L.-V., Heinrich, K., Wanner, J., Janiesch, C., 2023. Stop ordering machine learning algorithms by their explainability! A user-centered investigation of performance and explainability. Int J Inf Manage 69, 102538. https://doi.org/10.1016/j.ijinfomgt.2022.102538.

Herrmann, T., 2024. Comparing socio-technical design principles with guidelines for Human-centered AI. In: International Conference on Human-Computer Interaction.

Hoffman, R.R., Johnson, M., 2019. The quest for alternatives to "levels of automation" and "task allocation. In: Mouloua, M., Hancock, P.A. (Eds.), Human Performance in Automated and Autonomous Systems. CRC Press, pp. 43–68.

Hollnagel, E., Bye, A., 2000. Principles for modelling function allocation. Int J Hum Comput Stud 52 (2), 253–265. https://doi.org/10.1006/ijhc.1999.0288.

Holstein, K., De-Arteaga, M., Tumati, L., Cheng, Y., 2023. Toward supporting perceptual complementarity in Human-AI collaboration via reflection on unobservables. Proc, ACM Hum,-Comput, Interact 7 (CSCW1). https://doi.org/10.1145/3579628. Article 152.

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., Müller, H., 2019. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. 9 (4), e1312. https://doi.org/10.1002/widm.1312.

Holzinger, A., 2018a. From machine learning to explainable AI. In: 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA). Košice, Slovakia. https://doi.org/10.1109/DISA.2018.8490530.

Holzinger, A., 2018b. From machine learning to explainable AI. In: 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA).

Hoon, C., 2013. Meta-synthesis of qualitative case studies: an approach to theory building. Organ Res Methods 16 (4), 522–556. https://doi.org/10.1177/1094428113484969.

Hou, K., Hou, T., Cai, L., 2023. Exploring trust in Human–AI collaboration in the context of multiplayer online games. Systems 11, 217. https://doi.org/10.3390/systems11050217.

Hsieh, S.H., Lee, C.T., 2024. The AI humanness: how perceived personality builds trust and continuous usage intention. J. Prod. Brand Manag.

Hu, M., Zhang, G., Chong, L., Cagan, J., Goucher-Lambert, K., 2024. How being outvoted by AI teammates impacts Human-AI collaboration. Int. J. Hum.–Comput. Interact. 1–18. https://doi.org/10.1080/10447318.2024.2345980.

Hutchins, E., 1995. Cognition in the Wild. MIT Press. https://books.google.hu/books?id=CGIaNc3F1MgC.

Ibrahim, L., Ghassemi, M.M., Alhanai, T., 2023. Do explanations improve the quality of AI-assisted Human decisions? An algorithm-in-the-loop analysis of factual & counterfactual explanations. In: Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems.

Inkpen, K., Chappidi, S., Mallari, K., Nushi, B., Ramesh, D., Michelucci, P., Mandava, V., VepŘek, L.H., Quinn, G., 2023. Advancing Human-AI complementarity: the impact of user expertise and algorithmic tuning on joint decision making. ACM Trans. Comput.-Hum. Interact. 30 (5), 1–29. https://doi.org/10.1145/3534561.

Jain, R., Garg, N., Khera, S., 2022. Effective Human–AI work design for collaborative decision-making. Kybernetes. https://doi.org/10.1108/K-04-2022-0548.

Jakubik, J., Schoeffer, J., Hoge, V., Vössing, M., & Kühl, N. (2022). *An empirical evaluation of predicted outcomes as explanations in Human-AI decision-making*. https://doi.org/10.48550/arXiv.2208.04181.

Janssen, C.P., Donker, S.F., Brumby, D.P., Kun, A.L., 2019. History and future of Human-automation interaction. Int J Hum Comput Stud 131, 99–107. https://doi.org/10.1016/j.ijhcs.2019.05.006.

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

Jiang, J., Kahai, S., Yang, M., 2022. Who needs explanation and when? Juggling explainable AI and user epistemic uncertainty. Int J Hum Comput Stud 165, 102839. https://doi.org/10.1016/j.ijhcs.2022.102839.

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

Jussupow, E., Spohrer, K., Heinzl, A., Gawlitza, J., 2021. Augmenting medical diagnosis decisions? An investigation into physicians' Decision-making process with artificial intelligence. Inf. Syst. Res. 32. https://doi.org/10.1287/isre.2020.0980.

Kaber, D.B., Endsley, M.R., 1997. Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. Process saf. prog. 16 (3), 126–131. https://doi.org/10.1002/prs.680160304.

Kaber, D.B., Onal, E., Endsley, M.R., 2000. Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. Hum. factors ergon. manuf. serv. ind. 10 (4), 409–430. https://doi.org/10.1002/1520-6564(200023)10:4%3C409::AID-HFM4%3E3.0.CO;2-V.

Kahr, P., Rooks, G., Willemsen, M., & Snijders, C. (2023). *It seems smart, but It acts stupid: development of trust in AI advice in a repeated legal decision-making task*. https://doi.org/10.1145/3581641.3584058.

Kahr, P., Rooks, G., Willemsen, M., Snijders, C., 2024a. Understanding trust and reliance development in AI advice: assessing model accuracy, model explanations, and experiences from previous interactions. ACM Trans. Interact. Intell. Syst. https://doi.org/10.1145/3686164.

Kahr, P.K., Rooks, G., Snijders, C., Willemsen, M.C., 2024b. The trust recovery journey. The effect of timing of errors on the willingness to follow AI advice. In: Proceedings of the 29th International Conference on Intelligent User Interfaces.

Karran, A.J., Demazure, T., Hudon, A., Senecal, S., Léger, P.-M., 2022. Designing for confidence: the impact of visualizing artificial intelligence decisions. Front Neurosci 16, 883385. https://doi.org/10.3389/fnins.2022.883385.

Kelly, M., Kumar, A., Smyth, P., Steyvers, M., 2023. Capturing humans' Mental models of AI: an item response theory approach. In: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. Chicago, IL, USA. https://doi.org/10.1145/3593013.3594111.

Kim, S.S.Y., Watkins, E.A., Russakovsky, O., Fong, R., Monroy-Hernández, A., 2023a. "Help me Help the AI": understanding how explainability can support Human-AI interaction. In: 2023 CHI Conference on Human Factors in Computing Systems. Hamburg, Germany. https://doi.org/10.1145/3544548.3581001.

Kim, S.S.Y., Watkins, E.A., Russakovsky, O., Fong, R., Monroy-Hernández, A., 2023b. Humans, AI, and context: understanding end-users' Trust in a real-world computer vision application. In: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. Chicago, IL, USA. https://doi.org/10.1145/3593013.3593978.

Kim, S.S.Y., Liao, Q.V., Vorvoreanu, M., Ballard, S., Vaughan, J.W., 2024. I'm not sure, but...": *examining the impact of large language models' Uncertainty expression on user reliance and trust.* In: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency. Rio de Janeiro, Brazil. https://doi.org/10.1145/3630106.3658941.

Kong, H., Yin, Z., Chon, K., Yuan, Y., Yu, J., 2024. How does artificial intelligence (AI) enhance hospitality employee innovation? The roles of exploration, AI trust, and proactive personality. J. Hosp. Mark. Manag. 33 (3), 261–287. https://doi.org/10.1080/19368623.2023.2258116.

Kuang, E., Li, M., Fan, M., Shinohara, K., 2024. Enhancing ux evaluation through collaboration with conversational AI assistants: effects of proactive dialogue and timing. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. Honolulu, HI, USA. https://doi.org/10.1145/3613904.3642168.

Lai, V., Carton, S., Bhatnagar, R., Liao, V., Zhang, Y., & Tan, C. (2022). *Human-AI collaboration via conditional delegation: a case study of content moderation.* https://doi.org/10.1145/3491102.3501999.

Langley, A., 2021. What is "this" a case of? Generative theorizing for disruptive times. J. Manag. Inq. 30 (3), 251–258. https://doi.org/10.1177/10564926211016545.

Lauer, T., Wieland, S., 2021. Human-AI-collaboration in the context of information asymmetry – a behavioral analysis of demand forecasting. In: Ahram, T.Z., Karwowski, W., Kalra, J. (Eds.), Advances in Artificial Intelligence. Software and Systems Engineering, Cham.

Le, V.N.T., Kang, J., Oh, I.-S., Kim, J.-G., Yang, Y.-M., Lee, D.W., 2022. Effectiveness of Human–Artificial intelligence collaboration in cephalometric landmark detection. J Pers Med 12, 387. https://doi.org/10.3390/jpm12030387.

Lebovitz, S., Lifshitz-Assaf, H., Levina, N., 2022. To engage or not to engage with AI for critical judgments: how professionals deal with opacity when using AI for medical diagnosis. Organ. Sci. 33. https://doi.org/10.1287/orsc.2021.1549.

LeCun, Y., Boser, B.E., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W.E., Jackel, L.D., 1989. *Handwritten digit recognition with a back-propagation network* neural information processing systems. In: https://proceedings.neurips.cc/paper_files/paper/1989/hash/53c3bce66e43be4f209556518c2fcb54-Abstract.html.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep Learning. Nature 521 (7553), 436–444. https://doi.org/10.1038/nature14539.

Lee, M.H., Chew, C.J., 2023. Understanding the effect of counterfactual explanations on trust and reliance on AI for Human-AI collaborative clinical decision making. Proc. ACM Hum.-Comput. Interact. 7 (CSCW2), 1–22. https://doi.org/10.1145/3610218.

Lee, M., Siewiorek, D., Smailagic, A., Bernardino, A., & Bermúdez i Badia, S. (2021). *A Human-AI collaborative approach for clinical decision making on rehabilitation assessment.* https://doi.org/10.1145/3411764.3445472.

Lee, C., Kwon, H.Y., Cha, K.J., 2024. Human cognition for mitigating the paradox of AI explainability: a pilot study on Human gaze-based text highlighting. In: The International FLAIRS Conference Proceedings.

Leichtmann, B., Hinterreiter, A., Humer, C., Streit, M., Mara, M., 2023a. Explainable artificial intelligence improves Human decision-making: results from a mushroom picking experiment at a public art festival. Int. J. Hum.–Comput. Interact. 1–18. https://doi.org/10.1080/10447318.2023.2221605.

Leichtmann, B., Humer, C., Hinterreiter, A., Streit, M., Mara, M., 2023b. Effects of explainable artificial intelligence on trust and Human behavior in a high-risk decision task. Comput Hum. Behav 139, 107539. https://doi.org/10.1016/j.chb.2022.107539.

Lemaignan, S., Warnier, M., Sisbot, E.A., Clodic, A., Alami, R., 2017. Artificial cognition for social Human–Robot interaction: an implementation. Artif Intell 247, 45–69. https://doi.org/10.1016/j.artint.2016.07.002.

Lemus, H., Kumar, A., & Steyvers, M. (2023). How displaying AI confidence affects reliance and hybrid Human-AI performance. In. https://doi.org/10.3233/FAIA230087.

Leyer, M., Schneider, S., 2021. Decision augmentation and automation with artificial intelligence: threat or opportunity for managers? Bus Horiz 64 (5), 711–724. https://doi.org/10.1016/j.bushor.2021.02.026.

Li, J., Huang, J., Liu, J., Zheng, T., 2022. Human-AI cooperation: modes and their effects on attitudes. Telemat. Inform. 73, 101862. https://doi.org/10.1016/j.tele.2022.101862.

Li, Z., Lu, Z., Yin, M., 2023. Modeling Human trust and reliance in AI-assisted decision making: a Markovian approach. In: Proceedings of the AAAI Conference on Artificial Intelligence.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P.A., Clarke, M., Devereaux, P.J., Kleijnen, J., Moher, D., 2009. The Prisma Statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Ann, Intern, Med 151 (4), W–65.

Lin, H.E., McDonough, E.F., Lin, S.J., Lin, C.Y.Y., 2013. Managing the exploitation/exploration paradox: the role of a learning capability and innovation ambidexterity. J. Prod. Innov. Manag. 30 (2), 262–278. https://doi.org/10.1111/j.1540-5885.2012.00998.x.

Lindebaum, D., Vesa, M., Den Hond, F., 2020. Insights from "the machine stops" to better understand rational assumptions in algorithmic decision making and its implications for organizations. Acad. Manag. Rev. 45 (1), 247–263. https://doi.org/10.5465/amr.2018.0181.

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, H., Lai, V., Tan, C., 2021. Understanding the effect of out-of-distribution examples and interactive explanations on Human-AI decision making. Proc, ACM Hum,-Comput, Interact 5 (CSCW2). https://doi.org/10.1145/3479552. Article 408.

Liu, M.X., Sarkar, A., Negreanu, C., Zorn, B., Williams, J., Toronto, N., Gordon, A.D., 2023. "What it wants me to say": bridging the abstraction gap between end-user programmers and code-generating large language models. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. Hamburg, Germany. https://doi.org/10.1145/3544548.3580817.

Liu, B., 2021. In AI we trust? Effects of agency locus and transparency on uncertainty reduction in Human–AI interaction. J. Comput.-Mediat. Commun. 26 (6), 384–402. https://doi.org/10.1093/jcmc/zmab013.

Lobo, I., Koch, J., Renoux, J., Batina, I., & Prada, R. (2024, 2024). When should I lead or follow: understanding initiative levels in Human-AI collaborative gameplay.

Loske, D., Klumpp, M., 2021. Intelligent and efficient? An empirical analysis of Human–AI collaboration for truck drivers in retail logistics. Int. J. Logist. Manag. https://doi.org/10.1108/IJLM-03-2020-0149.

Lu, Z., Yin, M., 2021. Human reliance on machine learning models when performance feedback is limited. In: *Heuristics and Risks* Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Yokohama, Japan. https://doi.org/10.1145/3411764.3445562.

Lu, Z., Wang, D., Yin, M., 2024. Does more advice help? The effects of second opinions in AI-assisted decision making. Proc, ACM Hum,-Comput, Interact 8 (CSCW1). https://doi.org/10.1145/3653708. Article 217.

Luo, Y., Tseng, H.-H., Cui, S., Wei, L., Ten Haken, R.K., El Naqa, I., 2019. Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling. BJR| Open 1 (1), 20190021. https://doi.org/10.1259/bjro.20190021.

Ma, S., Lei, Y., Wang, X., Zheng, C., Shi, C., Yin, M., Ma, X., 2023. Who should I trust: AI or myself? Leveraging Human and AI correctness likelihood to promote appropriate trust in AI-assisted decision-making. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.

Ma, S., Wang, X., Lei, Y., Shi, C., Yin, M., Ma, X., 2024. "Are you really sure?" understanding the effects of Human self-confidence calibration in AI-assisted decision making. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. Honolulu, HI, USA. https://doi.org/10.1145/3613904.3642671.

Maier, T., Menold, J., McComb, C., 2022. The relationship between performance and trust in AI in E-finance. Front. Artif. Intell. 5, 891529. https://doi.org/10.3389/frai.2022.891529.

Manzey, D., Reichenbach, J., Onnasch, L., 2012. Human performance consequences of automated decision aids: the impact of degree of automation and system experience. J Cogn Eng Decis Mak 6 (1), 57–87. https://doi.org/10.1177/1555343411433844.

Meade, M.O., Richardson, W.S., 1997. Selecting and appraising studies for a systematic review. Ann, Intern, Med 127 (7), 531–537.

Milewski, A.E., Lewis, S.H., 1997. Delegating to software agents. Int J Hum Comput Stud 46 (4), 485–500. https://doi.org/10.1006/ijhc.1996.0100.

Miller, S.M., 2018. AI: augmentation, more so than automation. Asian Manag. Insights 5 (1), 1–12. https://ink.library.smu.edu.sg/ami/83.

Minsky, M.L., 1988. The Society of Mind. Simon & Schuster. https://www.google.co.uk/books/edition/Society_Of_Mind/bLDLllfRpdkC.

Montazeri, M., Montazeri, M., Montazeri, M., Beigzadeh, A., 2016. Machine learning models in breast cancer survival prediction. Technol. Health Care 24 (1), 31–42. https://doi.org/10.3233/THC-151071.

Moravec, H., 1988. Mind Children: The Future of Robot and Human Intelligence. Harvard University Press. https://www.google.co.uk/books/edition/Mind_Children/56mb7XuSx3QC.

Morrison, K., Shin, D., Holstein, K., Perer, A., 2023. Evaluating the impact of Human explanation strategies on Human-AI visual decision-making. Proc. ACM Hum.-Comput. Interact. 7 (CSCW1), 1–37. https://doi.org/10.1145/3579481.

Morrison, K., Spitzer, P., Turri, V., Feng, M., Kühl, N., Perer, A., 2024. The impact of imperfect Xai on Human-AI decision-making. Proc. ACM Hum.-Comput. Interact. 8 (CSCW1), 1–39. https://doi.org/10.1145/3579481.

Mozannar, H., Sontag, D.A., 2020. Consistent estimators for learning to defer to an expert. Int. Conf. Mach. Learn.

Mozannar, H., Lee, J., Wei, D., Sattigeri, P., Das, S., & Sontag, D. (2023). *Effective Human-AI teams via learned natural language rules and onboarding* NeurIPS 2023: advances in neural information processing systems 36, New Orleans, LA. https://doi.org/10.48550/arXiv.2311.01007.

Mucha, H., Robert, S., Breitschwerdt, R., Fellmann, M., 2021. Interfaces for explanations in Human-AI interaction: proposing a design evaluation approach. In: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems.

Muijlwijk, H., Willemsen, M.C., Smyth, B., & Ijsselsteijn, W.A. (2024). Benefits of Human-AI interaction for expert users interacting with prediction models: a study on marathon running.

Nabizadeh Rafsanjani, H., Nabizadeh, A.H., 2023. Towards Human-centered artificial intelligence (AI) in architecture, Engineering, and Construction (Aec) industry. Comput. Hum. Behav. Rep. 11, 100319. https://doi.org/10.1016/j.chbr.2023.100319.

Nascimento, N., Alencar, P., & Cowan, D. (2023). *Artificial Intelligence* versus *software engineers: an evidence-based assessment focusing on non-functional requirements*. https://doi.org/10.21203/rs.3.rs-3126005/v1.

Newell, A., Simon, H., 1956. The logic theory machine–a complex information processing system. IRE Trans. inf. theory 2 (3), 61–79. https://doi.org/10.1109/TIT.1956.1056797.

Niraula, D., Cuneo, K.C., Dinov, I.D., Gonzalez, B.D., Jamaluddin, J.B., Jin, J.J., Luo, Y., Matuszak, M.M., Ten Haken, R.K., Bryant, A.K., 2024. Intricacies of Human-AI interaction in dynamic decision-making for precision oncology: a case study in response-adaptive radiotherapy. medRxiv. https://doi.org/10.1101/2024.04.27.24306434.

O'Shea, K., & Nash, R. (2015). *An introduction to convolutional neural networks* (ArXiv e-prints). https://doi.org/10.48550/arXiv.1511.08458.

Okamura, K., Yamada, S., 2020a. Adaptive Trust Calibration for Human-AI collaboration. PLOS ONE 15 (2), e0229132. https://doi.org/10.1371/journal.pone.0229132.

Okamura, K., Yamada, S., 2020b. Empirical evaluations of framework for adaptive trust calibration in Human-AI cooperation. IEEE Access 8, 220335–220351. https://doi.org/10.1109/ACCESS.2020.3042556.

Olekanma, O., Dörfler, V., Shafti, F., 2022. Stepping into the participants' Shoes: the trans-positional cognition approach (Tpca). Int J Qual Methods 21, 1–15. https://doi.org/10.1177/16094069211072413.

Organizing Committee of the Madrid 2017 Critical Care Datathon, Reiz, A.N., Sagasti, F. M., González, M.Á., Malpica, A.B., Benítez, J.C.M., Cabrera, M.N., del Pino Ramírez, Á., Perdomo, J.M.G., Alonso, J.P., 2019. Big Data and machine learning in critical care: opportunities for collaborative research. Med. intensiva 43 (1), 52–57. https://doi.org/10.1016/j.medin.2018.06.002.

Ou, C., Mayer, S., & Butz, A.M. (2023, 2023). The impact of expertise in the loop for exploring machine rationality.

Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., Falco, G., Fiore, S.M., Garibay, I., Grieman, K., 2023. Six Human-centered artificial intelligence grand challenges. Int. J. Hum.–Comput. Interact. 39 (3), 391–437. https://doi.org/10.1080/10447318.2022.2153320.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., 2021. The Prisma 2020 statement: an updated guideline for reporting systematic reviews. bmj 372.

Pandya, R., Huang, S.H., Hadfield-Menell, D., & Dragan, A.D. (2019, 2019). Human-AI learning performance in multi-armed bandits.

Papachristos, E., Johansen, P.S., Jacobsen, R.M., Bysted, L.B.L., Skov, M.B., 2021. How do people perceive the role of AI in Human-AI collaboration to solve everyday tasks?. In: CHI Greece 2021: 1st International Conference of the ACM Greek SIGCHI Chapter. Greece. Online (Athens, Greece). https://doi.org/10.1145/3489410.3489420.

Papenkordt, J., 2024. Navigating transparency: the influence of on-demand explanations on non-expert user interaction with AI. In: Degen, H., Ntoa, S. (Eds.), Artificial Intelligence in HCI. Cham.

Parasuraman, R., Mouloua, M., 2018. Automation and Human Performance: Theory and Applications. Routledge.

Parasuraman, R., Sheridan, T.B., Wickens, C.D., 2000. A model for types and levels of Human interaction with automation. IEEE Trans. Syst. Man Cybern. - A: Syst. Hum. 30 (3), 286–297. https://doi.org/10.1109/3468.844354.

Park, S., Kim, H.K., Song, Y., Bang, S., Kim, J., Park, J., Park, J., 2022. Impact of expectation and performance on the user experience of AI systems. ICIC Int. 学会 https://doi.org/10.24507/icicelb.13.01.73.

Pataranutaporn, P., Liu, R., Finn, E., Maes, P., 2023. Influencing Human–AI interaction by priming beliefs about AI can increase perceived trustworthiness, empathy and effectiveness. Nat. Mach. Intell. 5 (10), 1076–1086. https://doi.org/10.1038/s42256-023-00720-7.

Peitgen, H.-O., Jürgens, H., Saupe, D., 1992. Chaos and Fractals: New Frontiers of Science. /2012. Springer. https://books.google.co.uk/books?id=jVpS_u0Lg4gC.

Peng, A., Nushi, B., Kiciman, E., Inkpen, K., & Kamar, E. (2022, 2022). Investigations of performance and bias in Human-AI teamwork in hiring.

Petticrew, M., 2001. Systematic reviews from astronomy to zoology: myths and misconceptions. Bmj 322 (7278), 98–101.

Pinski, M., Adam, M., & Benlian, A. (2023, 2023). AI knowledge: improving AI delegation through Human enablement.

Pisoni, G., D-R, N., Gijlers, H., Tonolli, L., 2021. Human-centered artificial intelligence for designing accessible cultural heritage. Appl. Sci. 11 (2), 870. https://doi.org/10.3390/app11020870.

Ployhart, R.E., Bartunek, J.M., 2019. Editors' Comments: there is nothing so theoretical as good practice—A call for phenomenal theory. Acad. Manag. Rev. 44 (3). https://doi.org/10.5465/amr.2019.0087.

Point, S., Fendt, J., Jonsen, K., 2017. Qualitative Inquiry in management: methodological dilemmas and concerns in meta-analysis. Eur. Manag. Rev. 14 (2), 185–204. https://doi.org/10.1111/emre.12097.

Press, L., 1971. Toward balanced man-Machine Systems. Int J Man Mach Stud 3 (1), 61–73. https://doi.org/10.1016/S0020-7373(71)80021-4.

Qian, C., & Wexler, J. (2024, 2024). Take it, leave it, or fix it: measuring productivity and trust in Human-AI collaboration.

Raisch, S., Krakowski, S., 2021. Artificial Intelligence and management: the automation–Augmentation paradox. Acad. Manag. Rev. 46 (1), 192–210. https://doi.org/10.5465/amr.2018.0072.

Rapp, A., Boldi, A., Curti, L., Perrucci, A., Simeoni, R., 2023. Collaborating with a text-based chatbot: an exploration of real-world collaboration strategies enacted during Human-chatbot interactions. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3544548.3580995. <conf-loc>, <city>Hamburg</city>, <country>Germany</country>, </conf-loc>.

Rastogi, C., Zhang, Y., Wei, D., Varshney, K.R., Dhurandhar, A., Tomsett, R., 2022. Deciding fast and slow: the role of cognitive biases in AI-assisted decision-making. Proc. ACM Hum.-Comput. Interact. 6 (CSCW1), 1–22. https://doi.org/10.1145/3512930.

Rechkemmer, A., Yin, M., 2022, 2022. In: When Confidence Meets Accuracy: Exploring the Effects of Multiple Performance Indicators on Trust in Machine Learning Models.

Reverberi, C., Rigon, T., Solari, A., Hassan, C., Cherubini, P., Antonelli, G., Awadie, H., Bernhofer, S., Carballal, S., Dinis-Ribeiro, M., Fernández-Clotett, A., Esparrach, G.F., Gralnek, I., Higasa, Y., Hirabayashi, T., Hirai, T., Iwatate, M., Kawano, M., Mader, M., Maieron, A., Mattes, S., Nakai, T., Ordas, I., Ortigão, R., Zúñiga, O.O., Pellisé, M., Pinto, C., Riedl, F., Sánchez, A., Steiner, E., Tanaka, Y., Cherubini, A., Group, G.I.G.C.S., 2022. Experimental evidence of effective Human–AI collaboration in medical decision-making. Sci Rep 12 (1), 14952. https://doi.org/10.1038/s41598-022-18751-2.

Revilla, E., Saenz, M.J., Seifert, M., Ma, Y., 2023. Human–Artificial intelligence collaboration in prediction: a field experiment in the retail industry. J. Manag. Inf. Syst. 40 (4), 1071–1098. https://doi.org/10.1080/07421222.2023.2267317.

Richens, J.G., Lee, C.M., Johri, S., 2020. Improving the accuracy of medical diagnosis with causal machine learning. Nat Commun 11 (1), 3923. https://doi.org/10.1038/s41467-020-17419-7.

Riedl, M.O., 2019. Human-centered artificial intelligence and machine learning. Hum. Behav. Emerg. Technol. 1 (1), 33–36.

Rizzi, D.A., 1994. Causal reasoning and the diagnostic process. Theor Med 15 (3), 315–333. https://doi.org/10.1007/BF01313345.

Robbemond, V., Inel, O., & Gadiraju, U. (2022, 2022). Understanding the role of explanation modality in AI-assisted decision-making.

Rondonotti, E., Hassan, C., Tamanini, G., Antonelli, G., Andrisani, G., Leonetti, G., Paggi, S., Amato, A., Scardino, G., Di Paolo, D., 2023. Artificial intelligence-assisted optical diagnosis for the resect-and-discard strategy in clinical practice: the Artificial intelligence bli characterization (Abc) study. Endoscopy 55 (01), 14–22. https://doi.org/10.1055/a-1852-0330.

Ruissalo, J., 2024. Transition to Human-AI work: shifts in routines' Dynamics and the implications for roles in knowledge-intensive work. https://hdl.handle.net/10125/107085.

Russell, S., Norvig, P., 2020. Artificial Intelligence: A Modern Approach, 4th ed. Pearson Education http://aima.cs.berkeley.edu/.

Salikutluk, V., Koert, D., Jäkel, F., 2023. Interacting with large language models: a case study on AI-aided brainstorming for guesstimation problems. Hhai 2023: Augmenting Human Intellect. IOS Press, pp. 153–167. https://doi.org/10.3233/FAIA230081.

Salikutluk, V., Schöpper, J., Herbert, F., Scheuermann, K., Frodl, E., Balfanz, D., Jäkel, F., Koert, D., 2024. An evaluation of situational autonomy for Human-AI collaboration in a shared workspace setting. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. Honolulu, HI, USA. https://doi.org/10.1145/3613904.3642564.

Salimzadeh, S., Gadiraju, U., 2024. When in doubt! understanding the role of task characteristics on peer decision-making with AI assistance. In: Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization.

Salimzadeh, S., He, G., & Gadiraju, U. (2023). *A missing piece in the puzzle: considering the role of task complexity in Human-AI decision making*. https://doi.org/10.1145/3565472.3592959.

Salimzadeh, S., He, G., Gadiraju, U., 2024. *Dealing with uncertainty: understanding the impact of prognostic* versus *diagnostic tasks on trust and reliance in Human-AI decision making*. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. Honolulu, HI, USA. https://doi.org/10.1145/3613904.3641905.

Sandelowski, M., Docherty, S., Emden, C., 1997. Qualitative metasynthesis: issues and techniques. Res Nurs Health 20 (4), 365–371. https://doi.org/10.1002/(SICI)1098-240X(199708)20:4<365::AID−NUR9>3.0.CO;2-E.

Sankaran, G., Palomino, M.A., Knahl, M., Siestrup, G., 2022. A modeling approach for measuring the performance of a Human-AI collaborative process. Appl. Sci. 12 (22).

Schaap, G., Van de Sande, Y., & Schraffenberger, H. (2024, 2024). Outperformed by AI: interacting with superhuman AI changes the way we perceive ourselves.

Schemmer, M., Hemmer, P., Nitsche, M., Kühl, N., & Vössing, M. (2022, 2022). A meta-analysis of the utility of explainable artificial intelligence in Human-AI decision-making.

Schemmer, M., Bartos, A., Spitzer, P., Hemmer, P., Kuhl, N., Liebschner, J., & Satzger, G. (2023a). Towards effective Human-AI decision-making: the role of Human learning in appropriate reliance on AI advice. *ArXiv, abs/2310.02108*. https://doi.org/10.48550/arXiv.2310.02108.

Schemmer, M., Kuehl, N., Benz, C., Bartos, A., & Satzger, G. (2023b, 2023). Appropriate reliance on AI advice: conceptualization and the effect of explanations.

Schreckenghost, D., Milam, T., Billman, D., 2014. Human performance with procedure automation to manage spacecraft systems. In: 2014 IEEE Aerospace Conference.

Schreibelmayr, S., Moradbakhti, L., Mara, M., 2023. First impressions of a financial AI assistant: differences between high trust and low trust users. Front. Artif. Intell. 6, 1241290. https://doi.org/10.3389/frai.2023.1241290.

Shaikh, S.J., Cruz, I.F., 2023. AI in Human teams: effects on technology use, members' Interactions, and creative performance under time scarcity. AI Soc 38 (4), 1587–1600. https://doi.org/10.1007/s00146-021-01335-5.
Shen, H., Liao, K., Liao, Z., Doornberg, J., Qiao, M., Van Den Hengel, A., & Verjans, J.W. (2021, 2021). Human-AI interactive and continuous sensemaking: a case study of image classification using scribble attention maps.
Sheridan, T.B., Maschinenbauingenieur, K., 2002. *Humans and Automation: System Design and Research Issues* (Vol. 280). Human Factors and Ergonomics Society, Santa Monica, CA.
Sheridan, T., Parasuraman, R., 2005. Human-automation interaction. Rev. Hum. Factors Ergon. 1, 89–129. https://doi.org/10.1518/155723405783703082.
Sheridan, T.B., & Verplank, W.L. (1978). *Human and computer control of undersea teleoperators*). https://doi.org/10.21236/ADA057655.
Sheridan, T.B., 1992. Telerobotics, Automation, and Human Supervisory Control. MIT press.
Sheridan, T.B., 2000. Function allocation: algorithm, alchemy or apostasy? Int J Hum Comput Stud 52 (2), 203–216. https://doi.org/10.1006/ijhc.1999.0285.
Shin, D., 2021. The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI. Int J Hum Comput Stud 146, 102551. https://doi.org/10.1016/j.ijhcs.2020.102551.
Shneiderman, B., 2020a. Design lessons from AI's two grand goals: human emulation and useful applications. IEEE Trans. Technol. Soc. 1 (2), 73–82. https://doi.org/10.1109/TTS.2020.2992669.
Shneiderman, B., 2020b. Human-centered artificial intelligence: reliable, safe & trustworthy. Int. J. Hum.–Comput. Interact. 36 (6), 495–504. https://doi.org/10.1080/10447318.2020.1741118.
Shneiderman, B., 2020c. Human-centered artificial intelligence: three fresh ideas. AIS Trans. Hum.-Comput. Interact. 12 (3), 109–124. https://doi.org/10.17705/1thci.00131.
Sivaraman, V., Bukowski, L.A., Levin, J., Kahn, J.M., Perer, A., 2023. Ignore, trust, or negotiate: understanding clinician acceptance of AI-based treatment recommendations in health care. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.
Sowa, K., Przegalinska, A., Ciechanowski, L., 2021. Cobots in knowledge work: human – AI collaboration in managerial professions. J Bus Res 125, 135–142. https://doi.org/10.1016/j.jbusres.2020.11.038.
Spillner, L., Ringe, R., Porzel, R., & Malaka, R. (2023). "My, My, how can I resist you?"-examining user reactions to bogus explanations of AI prediction.
Steyvers, M., Kumar, A., 2024. Three challenges for AI-assisted decision-making. Perspect. Psychol. Sci. 19 (5), 722–734. https://doi.org/10.1177/17456916231181102.
Subramanian, H.V., Canfield, C., Shank, D.B., Kinnison, M., 2023. Combining uncertainty information with AI recommendations supports calibration with domain knowledge. J Risk Res 26 (10), 1137–1152. https://doi.org/10.1080/13669877.2023.2259406.
Suffian, M., 2024. Explainable AI Assisted Decision-Making And human Behaviour, 2024. Computing, Internet of Things and Data Analytics, Cham.
Swaroop, S., Buçinca, Z., Gajos, K.Z., & Doshi-Velez, F. (2024, 2024). Accuracy-time tradeoffs in AI-assisted decision making under time pressure.
Swartout, W., Paris, C., Moore, J., 1991. Explanations in knowledge systems: design for explainable expert systems. IEEE Expert 6 (3), 58–64. https://doi.org/10.1109/64.87686.
Taesiri, M.R., Nguyen, G., Nguyen, A., 2022. Visual correspondence-based explanations improve AI robustness and Human-AI team accuracy. Adv Neural Inf Process Syst 35, 34287–34301.
Taudien, A., Fügener, A., Gupta, A., Ketter, W., 2022a. Calibrating users' Mental models for delegation to AI. In: ICIS 2022: 43rd International Conference on Information Systems: Digitization for the Next Generation. Copenhagen, Denmark. https://aisel.aisnet.org/icis2022/user_behaivor/user_behaivor/16.
Taudien, A., Fügener, A., Gupta, A., Ketter, W., 2022b. In: *The Effect of AI Advice on Human Confidence in Decision-Making* HICSS 2022: 55th Hawaii International Conference on System Sciences. Honolulu, HI. https://doi.org/10.24251/HICSS.2022.029.
Tejeda, H., Kumar, A., Smyth, P., Steyvers, M., 2022. AI-assisted decision-making: a cognitive modeling approach to infer latent reliance strategies. Comput. Brain Behav. 5 (4), 491–508. https://doi.org/10.1007/s42113-022-00157-y.
Tejeda Lemus, H., Kumar, A., Steyvers, M., 2022. An empirical investigation of reliance on AI-assistance in a noisy-image classification task. Hhai2022: Augmenting Human Intellect. IOS Press, pp. 225–237.
Templier, M., Paré, G., 2015. A framework for guiding and evaluating literature reviews. Commun. Assoc. Inf. Syst. 37 (1), 6.
Thomas, D.R., Lin, J., Gatz, E., Gurung, A., Gupta, S., Norberg, K., Fancsali, S.E., Aleven, V., Branstetter, L., Brunskill, E., Koedinger, K.R., 2024. In: *Improving Student Learning with Hybrid Human-AI Tutoring: A Three-Study Quasi-Experimental Investigation* LAK '24: 14th Learning Analytics and Knowledge Conference. Kyoto, Japan. https://doi.org/10.1145/3636555.3636896.
Timulak, L., 2009. Meta-analysis of qualitative studies: a tool for reviewing qualitative research findings in psychotherapy. Psychother. Res. 19 (4–5), 591–600. https://doi.org/10.1080/10503300802477989.
Timulak, L., 2013. Qualitative meta-analysis. In: Flick, U. (Ed.), The Sage Handbook of Qualitative Data Analysis. SAGE Publications, pp. 481–495.
Trist, E.L., Bamforth, K.W., 1951. Some social and psychological consequences of the Longwall method of coal-getting:an examination of the psychological situation and defences of a work group in relation to the social structure and technological content of the work system. Hum. Relat. 4 (1), 3–38. https://doi.org/10.1177/001872675100400101.
Tschandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., Janda, M., Lallas, A., Longo, C., Malvehy, J., Paoli, J., Puig, S., Rosendahl, C., Soyer, H.P., Zalaudek, I., Kittler, H., 2020. Human–Computer collaboration for skin cancer recognition. Nat, Med 26 (8), 1229–1234. https://doi.org/10.1038/s41591-020-0942-0.
van der Zander, Q.E.W., Roumans, R., Kusters, C.H.J., Dehghani, N., Masclee, A.A.M., de With, P.H.N., van der Sommen, F., Snijders, C.C.P., Schoon, E.J., 2024. Appropriate trust in artificial intelligence for the optical diagnosis of colorectal polyps: the role of Human/artificial intelligence interaction. Gastrointest, Endosc. https://doi.org/10.1016/j.gie.2024.06.029.
Vas, R., 2016. Studio: Ontology-Centric Knowledge-Based System. In: Gábor, A., Kő, A. (Eds.), Corporate Knowledge Discovery and Organizational Learning: The Role, Importance, and Application of Semantic Business Process Management. Springer International Publishing, pp. 83–103. https://doi.org/10.1007/978-3-319-28917-5_4.
Vasconcelos, H., Jörke, M., Grunde-McLaughlin, M., Gerstenberg, T., Bernstein, M.S., Krishna, R., 2023. Explanations can reduce overreliance on AI systems during decision-making. Proc. ACM Hum.-Comput. Interact. 7 (CSCW1), 1–38. https://doi.org/10.1145/3579605.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv Neural Inf Process Syst 30. https://doi.org/10.48550/arXiv.1706.03762.
von Eschenbach, W., 2021. Transparency and the black box problem: why we do not trust AI. Philos Technol 34. https://doi.org/10.1007/s13347-021-00477-0.
von Richthofen, G., Ogolla, S., Send, H., 2022. Adopting AI in the context of knowledge work: empirical insights from German organizations. Information 13 (4). https://doi.org/10.3390/info13040199.
Vygotsky, L.S., 1978. Mind in Society: The Development of Higher Psychological Processes, (paperback ed.). Harvard University Press.
Vygotsky, L.S., 1979. Consciousness as a problem in the psychology of behavior. Sov. Psychol. 17 (4), 3–35. https://doi.org/10.2753/RPO1061-040517043.
Walter, M.C., Broder, H.R., Förster, M., 2023. Boosting benefits, offsetting obstacles—The impact of explanations on AI users' Task performance. In: 18th International Conference on Wirtschaftsinformatik. Paderborn, Germany. https://doi.org/10.1007/978-3-031-80122-8_7.
Wang, P., Ding, H., 2024. The rationality of explanation or Human capacity? Understanding the impact of explainable artificial intelligence on Human-AI trust and decision performance. Inf Process Manag 61 (4), 103732. https://doi.org/10.1016/j.ipm.2024.103732.
Wang, X., Yin, M., 2022. Effects of explanations in AI-assisted decision making: principles and comparisons. ACM Trans. Interact. Intell. Syst. 12 (4), 1–36. https://doi.org/10.1145/3519266.
Wang, X., Yin, M., 2023. Watch out for updates: understanding the effects of model explanation updates in AI-assisted decision making. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.
Wang, D., Khosla, A., Gargeya, R., Irshad, H., Beck, A., 2016. *Deep learning for identifying metastatic breast cancer*). http://arxiv.org/pdf/1606.05718.
Wang, D., Weisz, J.D., Muller, M., Ram, P., Geyer, W., Dugan, C., Tausczik, Y., Samulowitz, H., Gray, A., 2019. Human-AI collaboration in data science: exploring data scientists' Perceptions of automated AI. Proc. ACM Hum.-Comput. Interact. 3 (CSCW), 1–24. https://doi.org/10.1145/3359313.
Wang, L., Lin, Z.Q., Wong, A., 2020. Covid-net: a tailored deep convolutional neural network design for detection of Covid-19 cases from chest X-ray images. Sci Rep 10. https://doi.org/10.1038/s41598-020-76550-z. Article 19549.
Wang, D., Wang, L., Zhang, Z., Wang, D., Zhu, H., Gao, Y., Fan, X., Tian, F., 2021. Brilliant AI doctor. In: *Rural Clinics: Challenges in AI-Powered Clinical Decision Support System Deployment* CHI '21: 2021 CHI Conference on Human Factors in Computing Systems. Yokohama, Japan. https://doi.org/10.1145/3411764.3445432.
Wang, X., Lu, Z., Yin, M., 2022. Will you accept the AI recommendation?. In: *Predicting Human Behavior in AI-Assisted Decision Making* Proceedings of the ACM Web Conference 2022, Virtual Event. Lyon, France. https://doi.org/10.1145/3485447.3512240.
Wang, D.-Y., Ding, J., Sun, A.-L., Liu, S.-G., Jiang, D., Li, N., Yu, J.-K., 2023a. Artificial intelligence suppression as a strategy to mitigate Artificial intelligence automation bias. J. Am. Med. Inform. Assoc. 30 (10), 1684–1692. https://doi.org/10.1093/jamia/ocad118.
Wang, G., Lu, S., Wang, K., 2023b. Effect of AI decision speed on user adoption in Human-AI collaboration: the moderating role of historical decision quality. https://aisel.aisnet.org/amcis2023/sig_odis/sig_odis/22.
Wang, W., Gao, G., Agarwal, R., 2023c. Friend or foe? Teaming between artificial intelligence and workers with variation in experience. Manage Sci. https://doi.org/10.1287/mnsc.2021.00588.
Wang, C., Chen, Z., Cangelosi, A., & Liu, H. (2024a). On the utility of external agent intention predictor for Human-AI coordination. *arXiv preprint arXiv:2405.02229*. https://doi.org/10.48550/arXiv.2405.02229.
Wang, Y., Jiang, Y., Tang, J., Zhou, X., 2024b. Task Delegation from AI to Humans: The Impact of AI Transparency on Human Performance, 2024//. E-Business. New Challenges and Opportunities for Digital-Enabled Intelligent Future, Cham.
Weber, S., Kordyaka, B., Palombo, R., Siemon, D., Niehaves, B., 2023. Is a fool with a (N AI) tool still a fool? An empirical study of the creative quality of Human–AI collaboration. In: ACIS 2023: 34th Australasian Conference on Information Systems. https://aisel.aisnet.org/acis2023/132.
Weiss, S.M., Kulikowski, C.A., Amarel, S., Safir, A., 1978. A model-based method for computer-aided medical decision-making. Artif Intell 11 (1), 145–172. https://doi.org/10.1016/0004-3702(78)90015-2.

Westphal, M., Vössing, M., Satzger, G., Yom-Tov, G.B., Rafaeli, A., 2023. Decision control and explanations in Human-AI collaboration: improving user perceptions and compliance. Comput Hum. Behav 144, 107714. https://doi.org/10.1016/j.chb.2023.107714.

Wickens, C.D., Sebok, A., Li, H., Sarter, N., Gacy, A.M., 2015. Using modeling and simulation to predict operator performance and automation-induced complacency with robotic automation: a case study and empirical validation. Hum Factors 57 (6), 959–975. https://doi.org/10.1177/0018720814566454.

Wilder, B., Horvitz, E., Kamar, E., 2021. In: *Learning to Complement Humans* Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. Yokohama, Yokohama, Japan. https://doi.org/10.48550/arXiv.2005.00582.

Willemink, M.J., Noël, P.B., 2019. The evolution of image reconstruction for ct—From filtered back projection to artificial intelligence. Eur Radiol 29 (5), 2185–2195. https://doi.org/10.1007/s00330-018-5810-7.

Wilson, H.J., Daugherty, P.R., 2018. Collaborative intelligence: humans and AI are joining forces [Article]. Harv Bus Rev 96 (4), 114–123. https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces.

Wu, T., Terry, M., Cai, C.J., 2022. AI chains: transparent and controllable Human-AI interaction by chaining large language model prompts. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. New Orleans, LA, USA. https://doi.org/10.1145/3491102.3517582.

Wuest, T., Weimer, D., Irgens, C., Thoben, K.-D., 2016. Machine learning in manufacturing: advantages, challenges, and applications. Prod Manuf Res 4 (1), 23–45. https://doi.org/10.1080/21693277.2016.1192517.

Xu, Y., Zhu, J.-Y., Eric, I., Chang, C., Lai, M., Tu, Z., 2014. Weakly supervised histopathology cancer image segmentation and classification. Med Image Anal 18 (3), 591–604. https://doi.org/10.1016/j.media.2014.01.010.

Xu, B., Song, X., Cai, Z., Lim, E., Tan, C.-W., & Yu, J. (2020). Artificial Intelligence or augmented intelligence: a case study of Human-AI collaboration in operational decision making.

Xu, C., Lien, K.-C., & Höllerer, T. (2023, 2023). Comparing zealous and restrained AI recommendations in a real-world Human-AI collaboration task.

Yang, S.J.H., Ogata, H., Matsui, T., Chen, N.-S., 2021. Human-centered artificial intelligence in education: seeing the invisible through the visible. Comput. Educ.: Artif. Intell. 2, 100008. https://doi.org/10.1016/j.caeai.2021.100008.

Yao, D., Holopainen, J., Laukkanen, T., 2024. Human-AI interaction – Is it trust or emotions that mediates behavioral intensions?. In: HICSS 57: Hawaii International Conference on System Sciences. Honolulu, HI. https://doi.org/10.24251/HICSS.2024.181.

Yildirim, N., Pushkarna, M., Goyal, N., Wattenberg, M., & Viégas, F. (2023, 2023). Investigating how practitioners use Human-AI guidelines: a case study on the people + AI guidebook.

Zahedi, Z., & Kambhampati, S. (2021). Human-AI symbiosis: a survey of current approaches. *arXiv preprint arXiv:2103.09990*. https://doi.org/10.48550/arXiv.2103.09990.

Zhang, Y., Liao, Q.V., & Bellamy, R.K.E. (2020, 2020). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making.

Zhang, Q., Lee, M.L., & Carter, S. (2022). You complete me: human-AI teams and complementary expertise.

Zhang, R., Flathmann, C., Musick, G., Schelble, B., McNeese, N.J., Knijnenburg, B., Duan, W., 2023. I know this looks bad, but I can explain: understanding when AI should explain actions in Human-AI teams. ACM Trans. Interact. Intell. Syst. https://doi.org/10.1145/3635474.

Zhang, S., Yu, J., Xu, X., Yin, C., Lu, Y., Yao, B., Tory, M., Padilla, L.M., Caterino, J., Zhang, P., 2024. Rethinking Human-AI collaboration in complex medical decision making: a case study in sepsis diagnosis. In: Proceedings of the CHI Conference on Human Factors in Computing Systems.

Zhu, Q., Wang, D., Ma, S., Wang, A.Y., Chen, Z., Khurana, U., & Ma, X. (2024, 2024). Towards feature engineering with Human and AI's knowledge: understanding data science practitioners' Perceptions in Human&AI-assisted feature engineering design.