SIGHCI 2022 Proceedings

Special Interest Group on Human-Computer Interaction

12-12-2022

THE NEW DREAM TEA(I)M? Rethinking Human-AI Collaboration based on Human Teamwork

Marvin Braun
University of Göttingen, marvin.braun@uni-goettingen.de

Maike Greve
University of Göttingen, maike.greve@uni-goettingen.de

Lutz M. Kolbe
University of Göttingen, lkolbe@uni-goettingen.de

THE NEW DREAM TEA(I)M? Rethinking Human-AI Collaboration based on Human Teamwork

Marvin Braun
Georg-August-Universität Göttingen
marvin.braun@uni-goettingen.de

Maike Greve
Georg-August-Universität Göttingen
maike.greve@uni-goettingen.de

Lutz M. Kolbe
Georg-August-Universität Göttingen
lkolbe@uni-goettingen.de

ABSTRACT

The continuing rise of artificial intelligence (AI) creates a new frontier of information systems that has the potential to change the future of work. Humans and AI are set to complete tasks as a team, using their complementary strengths. Previous research investigated several aspects of human-AI collaboration, such as the impact of human-AI teams on performance and how AI can be designed to complement the human teammate. However, experiments are suffering from a lack of comparability due to the unlimited configurations, which ultimately limits their implications. In this study, we develop an overarching framework for experiments on human-AI collaboration, using human teamwork as a theoretical lens. Our framework provides a novel, temporal structure for the research domain. Thereby, emerging topics can be clustered sequentially.

Keywords: human-AI collaboration, experiments, human teamwork, framework

MOTIVATION

Information technology continues to be revolutionized by the ongoing progress of artificial intelligence (AI). Through the adoption of AI, information systems (IS) gain unprecedented capabilities in information processing (McCorduck and Cfe, 2004; Berente et al., 2021). In this context, the future role of humans in workplaces is highly discussed due to the new capabilities of AI to conduct tasks that could previously only be completed by humans (Schuetz and Venkatesh, 2020). However, AI does come with its limitations. The technology produces unpredictable and unavoidable errors because most algorithms are based on statistical methods that will never reach an error-free accuracy. These circumstances create the urgent need for humans and AI to work in teams, which allows the utilization of complementary strengths of both entities (hybrid intelligence) (Dellermann et al., 2019; Rai et al., 2019). This complementary potential has also been recognized by businesses. Global Human Capital Trends states that 60% of organizations plan to use AI as a complementary factor rather than a replacement (Mallon et al., 2020).

With the changing role of AI, human-AI collaboration moves closer to human-human interaction. In general, human-AI collaboration includes coordination and planning activities, aside from the actual task (Seeber et al., 2020) and AI-based IS are not mere tools anymore. Thus, emerging research contributions increasingly investigate the impact of AI teammates in organizational settings (e.g., Rix & Hess, 2022). Despite the promising experiments conducted in human-AI collaboration, it is heavily limited by the lack of comparability of the experiments. Experiments are barely comparable due to the different tasks and mechanisms studied (Maedche et al., 2019; Fügener et al., 2021) at different points of time. Human teamwork literature provides extensively researched knowledge that can be utilized to structure **the temporal component** of human-AI collaboration and the missing comparability of tasks used in experiments. Bringing together the emerging growing research corpus of human-AI collaboration and the current lack of comparability, a research-driven framework that allows the classification of experiments is strongly needed. Thus, we derive the following research question:

> **RQ:** What are the temporal components of human-AI collaboration?

To answer this research question, we conduct a systematic literature review by following the approach of Webster and Watson (2002) with the theoretical lens of human teamwork (Rousseau et al., 2006). Thereby, we can map empiric works of human-AI collaboration into categories of teamwork and create an adapted framework of human-AI teamwork. Our contribution is two-fold. First, we synthesize the current state of human-AI collaboration and identify emerging topics. Second, we restructure the topics of human-AI collaboration along with the temporal concepts of human teamwork. The individual notions of human-AI collaboration will be put into a holistic perspective.

HUMAN-AI COLLABORATION RESEARCH

First, we will shortly define our understanding of AI and AI-based IS in organizational (workplace) contexts to provide a basis for the following concepts. While there is no fixed definition for AI (Berente et al., 2021), the term AI generally describes the research field that employs intelligent agents that aim to reach a specific goal (Russell and Norvig, 2009). In an organizational context, Berente et al. (2021) define AI as "the frontier of computational advancements that references human intelligence in addressing ever more complex decision-making problems" (Berente et al. 2021, p. 5).

In general, research on human-AI collaboration investigates how teams of humans and AI can conduct tasks and ultimately reach common goals (Rai et al., 2019). The research domain views AI-based IS more as digital assistants (Maedche et al., 2019) than tools that are used by humans (Seeber et al., 2020). Consequently, human-AI collaboration is also different from previous research on human-computer interaction, because it focuses on cooperative work instead of the presentation of information (Card et al., 2018). Furthermore, the importance of environmental constraints in human-AI collaboration has been emphasized in IS literature (Davenport and Kirby, 2016; Maedche et al., 2019; Fügener et al., 2021).

The human teamwork framework of Rousseau et al. (2006) provides the overarching structure for our literature review. Teams are the basic unit for structuring tasks and responsibilities in organizations (Devine et al., 1999; Rousseau et al., 2006) and consist of at least two individuals that cooperate to reach a common goal. This activity can be defined as 'teamwork' (Rousseau et al., 2006; McEwan et al., 2017). Teamwork can be divided into several chronologically ordered regulation phases that a team needs to complete to achieve high performance on a given task. Rousseau describes these phases as "preparation, execution, evaluation, and adjustment" (Rousseau et al. 2006, p. 550). During the **preparation** phase, teams need to plan how they want to conduct a specific task, and this includes the specification of a mission (analysis), the goal (specification), and concrete planning. Next, the team moves to the **execution** phase, i.e., they conduct the planned activities. During this phase, the team conducts "task-related collaborative behavior" (Rousseau et al. 2006, p. 551), i.e., coordination, cooperation, and information exchange. The **evaluation** phase includes work assessment behaviors, namely, monitoring performance and systems. Performance monitoring comprises of tracking a teammate's progress on the task and assuring that no mistakes are made during task completion. However, often if the task completion is not following the initial performance plan, hence the action regulation theory postulates that evaluation can be used to adjust the teamwork to complete the task (**team adjustment phase**). This can lead to either adapting the team's goal (i.e., teamwork behavior) or increasing efforts to reach the goal (i.e., non-teamwork behavior). Behavior adjustment means that team members help each other if they fail to complete their activities. Additionally, based on the evaluation phase, the team can provide feedback if failures are recognized to improve a teammate's performance (Rousseau et al., 2006).

METHODOLOGY

To synthesize the current state of human-AI collaboration in organizational contexts, we conducted a systematic literature review based on the proposed structure of Webster and Watson (2002). To put our literature review into perspective, we classify it with the provided taxonomy by Cooper (1988). Our literature review focuses on the research outcomes of human-AI collaboration in workplaces. The goal of our literature review is to integrate the different works that can be accounted for the topic and synthesize their content for future researchers. The review is organized in a conceptual form, i.e., we aggregate the researched topics of the different works. Our perspective is neutral. The audience is general and specialized scholars that are interested in human-AI augmentation in workplace contexts. Last, we conducted an exhaustive literature review with selection criteria (e.g., only works that were published after 2017).

RESULTS

The sample consists of 28 papers that include experiments in the context of human-AI collaboration. The main criterion for re-structuring the papers is the proposed structure of teamwork phases (preparation, execution, evaluation and adjustment). We combined the evaluation and adjustment phases because they were inseparable in all experiments. Seven experiments investigated human-AI collaboration during the preparation phase. Moreover, 21 experiments manipulated the execution phase (task conduction). Lastly, six papers investigated how teamwork can be evaluated and adjusted during or after task conduction. Some experiments included manipulations of multiple phases.

Initially, we classified the interaction type of the experiments as either "human-in-the-loop" or "computer-in-the-loop", depending on which entity supports and which entity is in charge. However, we found that this classification is not fruitful, because not all papers involve a clear hierarchy of humans and AI during. Thus, we decided to classify experiments following the proposed definitions of task assemblage and task augmentation (Rai et al., 2019). The interaction type was mostly determined as task augmentation (23). The most common form of augmentation was the AI supporting the human in decision-making. Five works investigated task assemblage. Experiments in this category involved more complex tasks humans and AI had to work on together as teammates. Last, we analyzed whether the experiments' dependent main variable is observable (e.g., team performance, decision making) or latent (e.g., trust, satisfaction). Most works investigated (team) performance as a central construct (24). Additionally, we identified the constructs of trust (6), and task delegation behavior (4), but also other variables such as cognitive load or team cognition (10).

As stated in the conceptual background, the preparation phase of teamwork aims to create a shared understanding of the task and goal and ultimately generates a plan for the team to reach the set goals. Works in the domain of teamwork demonstrate the importance of a proper preparation phase and its significant impact on team performance (Rousseau et al., 2006). We identify three major topics in this preparation phase. First, the major topic of studies investigating human-AI collaboration before conducting a task was to share information about the AI with the user to mainly create an understanding of the AI teammate's functioning (**team capability**) (Lai et al., 2020). Connected to this, several works researched how humans can be prepared to interact with AI (**team cognition**) (e.g., Weiler et al., 2021). In human-human teamwork, the process of sharing information about the competencies and weaknesses of team members is also prevalent and helps the team assess the performance of every member (Rousseau et al., 2006). Third, we identify papers investigating how tasks can be delegated to AI (**task delegation**).

The execution phase refers to all activities that happen during the actual teamwork. Human-human teamwork distinguishes between collaboration, coordination, and information exchange (about the tasks). In our sample, two types of human-AI teamwork emerged (task augmentation and task assemblage – Rai et al. 2019) that determine the **type of interaction** during the execution phase and impact the requirements of the AI and the work environment. Apart from the interaction type, a second research stream is identified. Information sharing during task execution is important for human teammates to understand and evaluate the behavior of their AI teammates. Together with the main measure of (team) performance, trust was often investigated in the context of providing explanations for the AI teammate's behavior (e.g., Buçinca et al., 2020). While prediction (15) is also the most utilized task type, experiments also involve recognition (7) (e.g., image detection), action tasks (6), and reasoning (4).

The type of interaction heavily impacted the human-AI teamwork during the task execution phase. The most dominant researched interaction form is decision support because the AI assists the human with the task (**task augmentation**). In these tasks, the AI recommends a decision that the human can either accept or reject (Shin et al., 2021). Most experiments focused on basic tasks that are domain-independent (e.g., Green and Chen, 2019; Fügener et al., 2021). Apart from the interaction form of task augmentation, a second research stream emerged. This research stream investigates how humans and AI cooperate with distinct tasks and a common goal (**task assemblage**). These works usually involved a game-like setting in the experiment (e.g., Musick et al., 2021; Schoonderwoerd et al., 2022) in which humans and AI have different roles. All studies point toward the importance of social factors and team composition during these situations

When AI is used as a system, humans tend to follow the AI's recommendations. There are attempts to change this teamwork behavior of humans by providing explanations for the recommendations (Bansal et al., 2021; Gajos and Mamykina, 2022). Across both interaction forms, explaining the AI's behavior was a crucial topic (**transparency and explanations**). A common approach in this context is to highlight important features that impacted the AI's decision. This approach highly varies depending on the information that needs to be processed (i.e., text or images). For instance, Bansal et al. (2021) highlighted the important textual passages that led to the specific recommendation. However, experiments investigating these explanations show that they are convincing the human teammate of not only the correct form but also at times the incorrect advice form and thus, such implementations can also be counterproductive (e.g., Bansal et al., 2021). The mechanism allows the user to assess the certainty of the algorithm about its decision and is often used in combination with an explanation. However, like explanations, the effect of confidence scores is beneficial for correct and malicious for incorrect recommendations (Buçinca et al., 2020; Bansal et al., 2021).

In traditional human-human teamwork, the phases of evaluation and adjustment are separated. In total, six papers combined the process of evaluating the performance and adjusting the teamwork. Thus, we combined the phases of evaluation and adjustment for this work. Three works investigated prediction tasks and three works involved action tasks. We identified two human-AI collaboration topics that combine multiple aspects of human-human teamwork. Multiple authors research the mechanism of monitoring the human-AI teams' performance and providing feedback based on it to the user (**performance monitoring and feedback**). The most common form of providing feedback is during the task after a subtask has been completed and the team's performance has been assessed (e.g., Bansal et al., 2021). Fügener et al. (2021a) researched the impact of monitored and communicated human and AI performance on delegation decisions.

However, they could not find any significant effect of feedback on delegation. In contrast, Yin et al. (2019) showed feedback to participants after the first half of the experiments (twenty decision tasks) and included the participant's performance, the AI's performance, and the agreement fraction in it. Their experiments indicate that feedback on the AI's performance can affect people's trust in the AI and is also altered by the observed AI performance of the human.

DISCUSSION

This study synthesizes the current state of human-AI collaboration and explores how existing research, and its experiments can be restructured by utilizing human teamwork as a theoretical lens. The main contribution of our literature analysis is an integrative work that structures emerging topics of human-AI collaboration in a **temporal framework** (preparation, execution and evaluation and adjustment) and includes essential components of human-AI collaboration (user, task, AI - (Rzepka and Berger, 2018). While this might seem trivial at first, the sequential nature of the framework allows to group and link different research topics and provides overarching rationales from human teamwork. Previously, human-AI collaboration topics were analyzed from an actor-based view (e.g., Rzepka & Berger, 2018). Through the analysis of the experiments from a temporal perspective, we find that there is a current focus on the execution phase of human-AI collaboration. Naturally, this is the most prominent phase of teamwork, however, it also shows that there is a crucial knowledge gap in the preparation and in the preparation and adjustment phase.

Our work sets up several foundations for the future of work and AI developers. As discussed in the introduction, AI has the potential to revolutionize IS in work contexts. On this note, our results are two-fold. First, human-AI teamwork was previously often depicted as a unidimensional construct (e.g., Rzepka and Berger, 2018). We extend this perspective by creating a sequential framework of human-AI teamwork. This newly created understanding can help developers, but also managers in organizations to improve human-AI teamwork in several ways. Our work underlines the importance of providing team cognition when working with smart systems that make unforeseen errors. Guidelines and classified examples showed a great improvement in team performance among human-AI teams. Developers can use this insight to create impactful tutorials for interacting with their AI-based IS. Organizations introducing AI-based IS can obtain real-world examples of errors produced by the AI using real data and train their employees to handle their AI. Moreover, our results suggest that several other mechanisms, such as delaying recommendations of AI, can be beneficial for team performance depending on the type of task (e.g., Park et al., 2019). Second, AI achieves human-like performance or even beats a human and a human-AI team in experimental settings. However, almost all experiments conducted involved simple decision tasks that neither have high complexity nor involve high-stake decisions. While we acknowledge the potential of AI to transform workplaces and be a partner of human teammates in organizational settings, AI is still usually only applied to a single task. Moreover, when experiments included more complex tasks, the AI was usually fictive. Therefore, the second practical implication of our work is that managers should focus on the implementation of AI for basic, narrow tasks, where the human stays in the lead and receives support from the AI-based IS.

REFERENCES

1. Bansal, G., T. Wu, J. Zhou, R. Fok, B. Nushi, E. Kamar, … D. Weld. (2021). Does the Whole Exceed its Parts? The Effect of AI Explanations on Complementary Team Performance. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–16. New York, NY, USA: ACM.
2. Berente, N., J. Recker and R. Santhanam. (2021). Managing Artificial Intelligence. *MIS Quarterly: Management Information Systems 45* (3), 1433–1450.
3. Buçinca, Z., P. Lin, K. Z. Gajos and E. L. Glassman. (2020). Proxy tasks and subjective measures can be misleading in evaluating explainable AI systems. In: *Proceedings of the 25th International Conference on Intelligent User Interfaces*, pp. 454–464. New York, NY, USA: ACM.
4. Card, S. K., T. P. Moran and A. Newell. (2018). *The Psychology of Human-Computer Interaction*. (S. K. Card, Ed.). CRC Press.
5. Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society 1* (1), 104–126.
6. Davenport, T. H. and J. Kirby. (2016). Just How Smart Are Smart Machines. *MIT Sloan Management Review 57*, 21–25.
7. Dellermann, D., P. Ebel, M. Söllner and J. M. Leimeister. (2019). Hybrid Intelligence. *Business & Information Systems Engineering 61* (5), 637–643.
8. Devine, D. J., L. D. Clayton, J. L. Philips, B. B. Dunford and S. B. Melner. (1999). Teams in Organizations. *Small Group Research 30* (6), 678–711.
9. Fügener, A., J. Grahl, A. Gupta and W. Ketter. (2021). Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation. *Information Systems*

Research.

10. Gajos, K. Z. and L. Mamykina. (2022). Do People Engage Cognitively with AI? Impact of AI Assistance on Incidental Learning. In: *27th International Conference on Intelligent User Interfaces*, pp. 794–806. New York, NY, USA: ACM.

11. Green, B. and Y. Chen. (2019). The Principles and Limits of Algorithm-in-the-Loop Decision Making. *Proceedings of the ACM on Human-Computer Interaction 3* (CSCW), 1–24.

12. Lai, V., H. Liu and C. Tan. (2020). 'Why is "Chicago" deceptive?' Towards Building Model-Driven Tutorials for Humans. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. New York, NY, USA: ACM.

13. Maedche, A., C. Legner, A. Benlian, B. Berger, H. Gimpel, T. Hess, … M. Söllner. (2019). AI-Based Digital Assistants. *Business & Information Systems Engineering 61* (4), 535–544.

14. Mallon, D., J. Harris, B. Denny, J. Schwartz and S. Elliott. (2020). Superteams : Putting AI in the group CAPITAL H Superteams : Putting AI in the group. Retrieved from https://www2.deloitte.com/us/en/insights/focus/human-capital-trends/2020/human-ai-collaboration.html

15. McCorduck, P. and C. Cfe. (2004). *Machines Who Think*. A K Peters/CRC Press.

16. McEwan, D., G. R. Ruissen, M. A. Eys, B. D. Zumbo and M. R. Beauchamp. (2017). The Effectiveness of Teamwork Training on Teamwork Behaviors and Team Performance: A Systematic Review and Meta-Analysis of Controlled Interventions. *PLOS ONE 12* (1), e0169604.

17. Musick, G., T. A. O'Neill, B. G. Schelble, N. J. McNeese and J. B. Henke. (2021). What Happens When Humans Believe Their Teammate is an AI? An Investigation into Humans Teaming with Autonomy. *Computers in Human Behavior 122*, 106852.

18. Park, J. S., R. Barber, A. Kirlik and K. Karahalios. (2019). A Slow Algorithm Improves Users' Assessments of the Algorithm's Accuracy. *Proceedings of the ACM on Human-Computer Interaction 3* (CSCW), 1–15.

19. Rai, A., P. Constantinides and S. Sarker. (2019). Next generation digital platforms : toward human-AI hybrids. *MIS Quarterly 44* (1), iii–ix.

20. Rix, J. and T. Hess. (2022). Hello, Mate! Insights from the Field on Leveraging Machine Teammates in Organizations. Pacific Asia Conference on Information Systems (PACIS) 2022.

21. Rousseau, V., C. Aubé and A. Savoie. (2006). Teamwork Behaviors. *Small Group Research 37* (5), 540–570.

22. Russell, S. and P. Norvig. (2009). *Artificial Intelligence: A Modern Approach*, 3rd Edition. USA: Prentice Hall Press.

23. Rzepka, C. and B. Berger. (2018). User Interaction with AI-enabled Systems: A systematic review of IS research. In: *International Conference on Information Systems 2018, ICIS 2018*. International Conference on Information Systems (ICIS).

24. Schoonderwoerd, T. A. J., E. M. van Zoelen, K. van den Bosch and M. A. Neerincx. (2022). Design patterns for human-AI co-learning: A wizard-of-Oz evaluation in an urban-search-and-rescue task. *International Journal of Human-Computer Studies 164*, 102831.

25. Schuetz, S. and V. Venkatesh. (2020). "Research Perspectives: The Rise of Human Machines: How Cognitive Computing Systems Challenge Assumptions of User-System Interaction ". *Journal of the Association for Information Systems* 460–482.

26. Seeber, I., L. Waizenegger, S. Seidel, S. Morana, I. Benbasat and P. B. Lowry. (2020). Collaborating with technology-based autonomous agents. *Internet Research 30* (1), 1–18.

27. Shin, W., J. Han and W. Rhee. (2021). AI-assistance for predictive maintenance of renewable energy systems. *Energy 221*, 119775.

28. Webster, J. and R. T. Watson. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly 26* (2), xiii–xxiii.

29. Weiler, S., C. Matt and T. Hess. (2021). Immunizing with information – Inoculation messages against conversational agents' response failures. *Electronic Markets*.

30. Yin, M., J. Wortman Vaughan and H. Wallach. (2019). Understanding the Effect of Accuracy on Trust in Machine Learning Models. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12. New York, NY, USA: ACM.