

# Empirical Impacts of Independent and Collaborative Training on Task Performance and Improvement in Human-AI Teams

Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2024, Vol. 68(1) 1447–1453  
Copyright © 2024 Human Factors and Ergonomics Society  
DOI: 10.1177/10711813241274425  
journals.sagepub.com/home/pro



Christopher Flathmann<sup>1</sup> , Beau G. Schelble<sup>2</sup> ,  
and Anna Galeano<sup>1</sup>

## Abstract

With improving AI technology, human-AI teams are becoming increasingly common in research. Within these teams, humans and AI can work collaboratively to complete shared tasks. However, continuing research efforts highlight that humans are ill-prepared to work in human-AI teams. As such, recent efforts have called for training to become a greater focus in the. This paper reports on an empirical in-person experiment that explored the impact of individual and collaborative task-focused training in human-AI teams. Participants were tasked with either training together or separately prior to collaborative working in a human-AI team. Results show that having humans train together prior to joining a human-AI team can negatively impact their performance and ability to improve at a task when they begin working in a human-AI team. As such, results suggest that human-AI teams need to identify ideal ways to collaboratively train humans on task-related skills in human-AI teams.

## Keywords

human-AI teaming, artificial intelligence, training, performance

## Introduction

As the integration of artificial intelligence (AI) technologies into social and complex environments progresses, the need to improve how humans understand such technologies increases. One such integration of AI technology where this need is paramount is that of human-AI teams. These teams consist of at least one human team member working alongside at least one artificial agent, each working within their own role toward a common goal (McNeese et al., 2023). Within these teams, high expectations ultimately lead to AI teammates falling short of ideals in social and task-based interaction (Zhang et al., 2021). These difficulties lead to frustration and misunderstanding when teaming with a given AI system. In turn, human performance in these teams can suffer (Bansal et al., 2019).

Importantly, this struggle is not entirely unique to human-AI teams, as human-human teams also require adaptation, adjustment, and learning. Traditionally, these teams have relied on training to prepare human teammates to complete a team's task (Salas et al., 2008a). When intelligently used, training can drastically improve the performance of individuals and teams in the short and long term (Salas et al., 2008b). Most notably, teams can leverage training to target and

improve skills related to task completion and general teamwork (Prichard & Ashleigh, 2007). For task competencies, procedural-type training can help explain tasks to teammates and highlight the key skills they need to develop prior to working in actual teams (Sawyer & Gray, 2016). On the other hand, teamwork-focused training, such as cross or perturbation training, can help improve core teaming skills like communication, backup behaviors, and coordination (Marks et al., 2002). Based on a team's deficiencies and needs, they can then determine the best training for them.

Unfortunately, despite training's continued effectiveness in human-human teams, the concept has stayed relatively dormant in human-AI teaming research. This gap has ultimately led multiple meta-works to identify training as one of the most important gaps that need to be solved by human-AI teaming research (Oswald et al., 2022). As it stands,

<sup>1</sup>Clemson University, Clemson, SC, USA

<sup>2</sup>The University of Tennessee, Knoxville, TN, USA

## Corresponding Author:

Christopher Flathmann, Clemson University, 116 McAdams Hall,  
Clemson, SC 29634-0001, USA.

Email: cflathm@clemson.edu

current human-AI teaming research has shown preliminary benefits from cross and perturbation training, which are teamwork-focused training methodologies (Nikolaïdis et al., 2015; Johnson et al., 2023). However, this type of training can be resource-intensive and may not fully train human teammates on task competency. Indeed, training research needs to ensure that teamwork-focused training is balanced with task-focused training, but the latter is under-explored in human-AI teams.

As such, additional research is needed to understand how to best incorporate task-oriented training in human-AI teams. Specifically, a particular question that needs to be addressed relates to whether humans in human-AI teams should train on their tasks independently or collaboratively with other humans. With collaborative training having seen success in human-only team research (Liang et al., 1995; Prichard et al., 2006), it is also necessary to examine such training in human-AI teaming through its own unique lens as it is such a departure from traditional human-only teaming (McNeese et al., 2023).

As such, the current research paper explores the topic of individual versus collaborative training on the performance of human-AI teams. Through the analysis of an empirical experiment of 32 two-person human-AI teams, this study explores the impact that task-focused training can have on the performance of humans in human-AI teams. Within this experiment, participants were either given the opportunity to train alone or together prior to working in their human-AI team. In doing so, this study addressed the following research questions, which are explored throughout this paper:

*RQ1: Do individual or collaborative types of training in human-AI teams affect the performance of human-AI teams over time?*

*RQ2: Does training type affect human teammates' ability to improve their performance in human-AI teams?*

## Background

Human-AI teamwork is becoming one of the most rapidly accelerating research domains within human factors and human-computer interaction spaces. At a base level, human-AI teaming simply refers to a teaming construct where humans work interdependently with autonomous AI technologies to complete shared goals (O'Neill et al., 2022). However, as the construct has evolved over time, what it means to be a successful and effective human-AI team has evolved. Indeed, while past explorations of AI technology focused on maximizing the performance and contribution of AI teammates, recent shifts in the domain have moved the conversation toward human and AI alignment and compatibility (Ji et al., 2023). In this vein, ongoing research has acknowledged that AI performance is only a small component of human-AI team efficiency, and human capabilities

and contributions need to be similarly prioritized (Bansal et al., 2019). Resulting of this shift, more recent efforts in human-AI teaming research have moved to identify how AI teammates can be engineered and designed to improve this alignment and compatibility. For example, ongoing research on explainable AI technology has helped highlight how AI teammates can provide additional information to improve the performance, trust, and situation awareness that human teammates form (Lyons et al., 2023). Similarly, research has noted that transparent interfaces can help further explain the complex calculations and analysis being performed by AI in real time, increasing human situation awareness (Endsley, 2023). In sum, research can and should continue to understand how the intelligent design of AI teammates can help shift human-AI teams toward greater alignment.

However, it is worth noting that alignment, like most teaming constructs, is bidirectional (Shen et al., 2024). Indeed, while designing AI teammates to improve their alignment with human goals and needs is essential, it's also worth exploring how humans can best adapt to the goals of human-AI teams and the design of AI teammates. Most recently, human-AI teaming research has noted that humans in human-AI teams are able and willing to observe and adapt to their AI teammates (Flathmann et al., 2024). Similarly, humans can bolster the capabilities and performance of their AI teammates by providing a greater deal of information transparency and communication (Chen et al., 2018). As such, researchers are aware that humans can directly change and adapt to benefit their human-AI team, yet this process may require guidance and training.

Historically, training has been one of the most important tools available to teams, as it provides them the ability to develop and cultivate important skills outside of an environment where performance and efficiency matter (Salas et al., 2008a). In turn, teaming research has developed several specific training methodologies that target different tasks and teaming skills. For example, team cross-training consists of having teammates swap roles and responsibilities, which provides them with a level of awareness and understanding of how their team can work interdependently (Marks et al., 2002). Similarly, perturbation training can target team skills related to situation awareness and backup behaviors, as random perturbations, or roadblocks, can be used to trigger and measure a team's ability to detect and respond to such roadblocks (Gorman et al., 2007). Lastly, on the task side, procedural training, which teaches individual teammates their individual responsibilities within a task, can be used to provide task-specific skills that enable the completion of a teammate's role (Gorman et al., 2007). Additionally, design modifications can be made to each of these training methodologies, which can help cater to the specific needs of a team. Most relevant to this study, task-focused training can be conducted individually or collaboratively (Liang et al., 1995). Individual training and learning can be ideal for the development of an individual's task-related skills; however, the



**Figure 1.** Rocket league screenshot.

interdependent nature of teams means that collaborative training could ensure the development of these skills consistent with the requirement to collaborate and coordinate (Ellington & Dierdorff, 2014). All in all, none of these methods or modifications are superior to one another, as they each have their strengths and weaknesses; rather, teams need to identify their specific need and the training methodology that corresponds to that need.

However, despite being a fundamental tool in the improvement of human teammates, training work is nascent in HATs. Within human-robot teams, training is a well-established concept, with repeated studies showing its effectiveness (Nikolaidis et al., 2015). However, human-AI teams have only seen initial explorations into perturbation training (Johnson et al., 2023), which also isolates human-AI team training research from team-oriented training. Indeed, while human-AI teaming experiments leverage task and procedural training within their experimental procedure (McNeese et al., 2018), research has yet to step back and explore the actual design of this training. Due to this gap and the overwhelming need for training in human-AI teams, recent meta-works have explicitly called for the creation, evaluation, and application of training in human-AI teaming research (McNeese et al., 2023; Oswald et al., 2022). Moving forward, HAT research needs to start to remedy these gaps. Further, these gaps need to be remedied with empirical work that tries to quantify and qualify the impacts, design, and challenges of training in human-AI teams.

## Methods

### Experimental Task

This study leveraged an in-person experiment in the popular Esports platform Rocket League. Rocket League is a team-based soccer Esport that centers around teams dribbling,

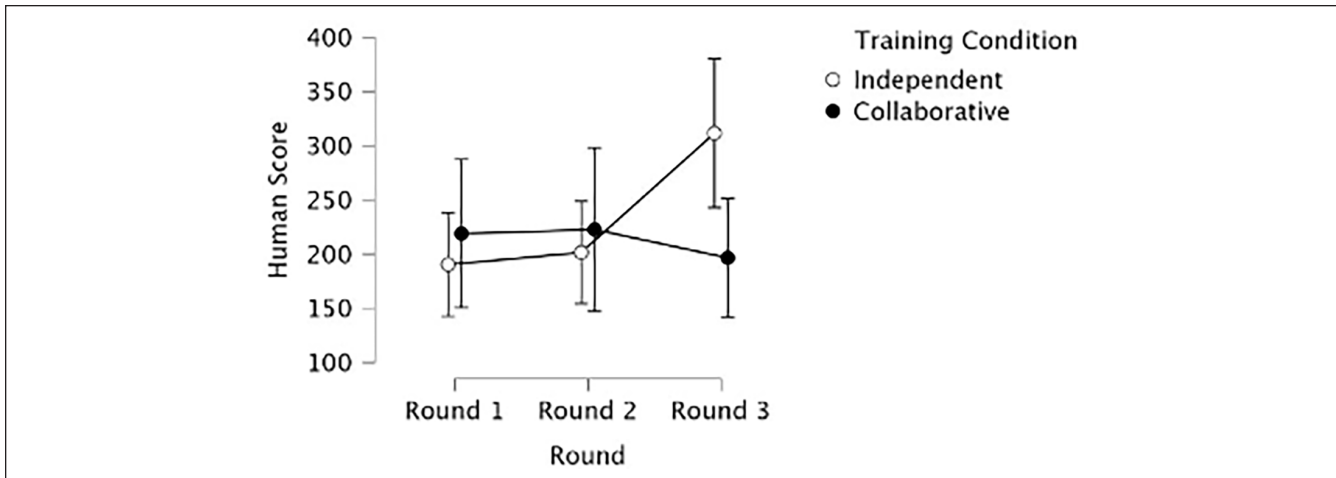
passing, and shooting a soccer ball while controlling cars (Pictured in Figure 1). In recent years, Rocket League has become a common platform for human-AI teaming research due to a growing repository of AI teammates that are validated in collaborative tournaments (see RLBot Community) (Flathmann et al., 2023; Verhoeven & Preuss, 2020). Within this environment, two participants were placed into a team with multiple AI teammates to form a human-AI team. These human-AI were tasked to score as many goals as possible on an opposing team of goalies.

### Experimental Design

This study leveraged a single between-subjects manipulation with two experimental conditions: individual and collaborative training. In the individual training condition, participants were tasked with completing multiple training games of Rocket League alone, without any other human or AI teammates. In turn, teams within this condition had both participants complete two individual training games separately prior to working together in a human-AI team. Alternatively, the collaborative training condition had both participants work together, without any AI teammates, to complete two training games. The length and purpose of these games were consistent across both conditions, and the only difference was whether participants completed these training games together or separately.

### Participants

In total, 32 teams were recruited to take part in this study, with two participants per team for a total of 64 participants. All participants were recruited via a university-maintained student subject pool. The average participant age was 18.34 ( $SD=0.65$ ) years old. Participants were compensated with course credit for participation in this experiment.



**Figure 2.** Graphical summary of the effect of round and training type on human teammate score.

### Experimental Procedure

After being recruited for this study, two participants arrived at a designated research lab. Once both arrived, participants were immediately provided with informed consent letters. Once these letters were read and agreed to, participants completed a few demographic questions. Once completed, participants were then given a brief tutorial on the Rocket League platform, which taught them aspects of the game itself. This tutorial was followed up with a hands-on tutorial and free-play session where participants were able to learn more about the platform and ask questions.

Once participants finished this session, the official procedural training began. In this training, participants were loaded into two 5-min practice games, where they completed a simulated game like their actual task. Based on their assigned manipulation, participants either completed both training games alone or separately.

Once finished with training, participants were placed on the same team with multiple other AI teammates. Within these human-AI teams, participants were then tasked with completing a series of three games in the Rocket League platform, which will be referred to as rounds. Participants were not allowed to talk with each other during either the training games or these three scored rounds. These rounds provided an isolated instance to measure team performance and interaction, and the series of rounds provided a means of exploring performance improvement. Breaks were provided in between each round so participants could rest and complete some additional surveys. Once all three rounds were completed, participants were dismissed and provided credit for their participation.

### Measures

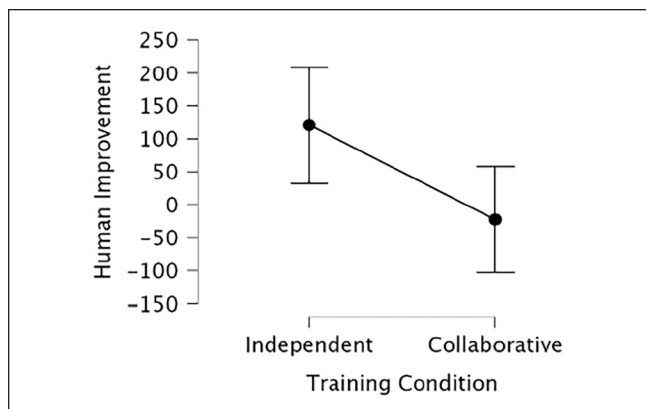
Measurement for this task was derived from the Rocket League scoring mechanism, as it provides a holistic score

based on participants' ability to move the ball, assist teammates, shoot the ball accurately, and score goals on the opposing team. This scoring mechanism has been used in prior HAT research as an accurate depiction of team and individual performance. With this metric, two dependent variables were calculated. First, human teammate performance was calculated for each human AI teammate for each round, and these performance scores were summed to create a human team score measure for each round. Additionally, the improvement participants saw from round 1 to round 3 was also calculated for each participant and summed for each team to create a human improvement metric.

### Results

For human teammates' summative score, the analysis used a  $2 \times 3$  mixed effects ANOVA with a within-subjects effect for the round (Round One, Round Two, Round Three) and a between-subjects effect for the team's training type (independent vs. collaborative; Figure 2). Results show a significant interaction effect between round and a team's training type ( $F [2, 60]=3.93, p=.025, \eta_p^2=.02$ ). An examination of the holm post hoc effects revealed that teammates who trained independently performed significantly better in the final round of the task (Round three) than teams that trained collaboratively. However, no significant differences existed between conditions in the first two rounds of performance.

For improvement, a  $1 \times 2$  ANOVA (statistically equivalent to an independent samples  $t$ -test) was conducted with a between-subjects effect for the team's training type (independent vs. collaborative; Figure 3). Results show a significant impact of a team's training condition on human teammates' improvement from Round 1 to Round 3 ( $F [1, 30]=6.53, p=.016, \eta_p^2=.18$ ). Specifically, human teammates who trained independently improved significantly more across all three rounds than those who trained collaboratively.



**Figure 3.** Graphical summary of the effect of training type on human teammate improvement across rounds.

These results highlight that participants who underwent task-focused training independently were significantly more capable of performing and improving their performance when working in a human-AI team. In turn, participants who trained together may have struggled to adapt their existing teamwork to accommodate AI teammates, ultimately worsening their performance. For task-focused training, having humans train independently may bolster individual performance in human-AI teams over the long term, as humans will be more adept at improving around both their human and AI teammates.

## Discussion

The immediate takeaway and contribution of this work is the potential negative impact of humans collaboratively training on tasks prior to entering a human-AI team despite this training showing effectiveness in human-human teamwork (Prichard et al., 2006). While prior research has continuously shown the importance of collaborative and team training to the long-term performance of teams (Salas et al., 2008a, b), the results of this study highlight a necessary but potentially overlooked consideration within human-AI teams. Collaborative training in its current form, with humans working together to complete simplified or modified tasks, may be insufficient for human-AI teams. Most likely, this ineffectiveness stems from the humans forming a strong shared understanding of each other as they learn to coordinate and collaborate without AI teammates (Schelble et al., 2022). While the training in question was task-focused, humans were still collaboratively developing team-related skills of coordination and collaboration through their interaction (Entin & Serfaty, 1999). As such, when these teams transition into a human-AI team, they are required to adjust their existing understanding of coordination and collaboration. Interestingly, this adjustment does not appear to harm the immediate performance of humans, as noted by an insignificant difference in round 1 and 2 performance. However, this adjustment impacts the improvement of human teammates over time.

As it stands, the overwhelming majority of teams within society are human. These teams are actively training and learning collaboration and coordination skills that are unique to human-human teams. With real-world teams beginning to contemplate the use of AI technology in the workplace (Farrow, 2022), these teams are likely going to find it difficult to adjust their existing collaboration and coordination skills to accommodate these AI systems. As such, human-AI team training research needs to be aware of the impact that collaborative training can have and work to create task-focused training that specifically helps existing workers learn and improve within human-AI teams.

Specifically, the results of this work suggest that an explicit division likely needs to be made between individual and collaborative task-focused training in HATs. Individual training will likely remain important for human-AI teams, as humans will need a way to cultivate their task-related skills in a safe and open environment. On the collaborative side, having humans collaborate is likely ineffective for human-AI teaming research. However, research should explore the integration of autonomous teammates in collaborative training. On the one hand, humans may find it beneficial to train with both human and AI teammates collaboratively; however, doing so may be overwhelming due to the novelty of AI teammates and their complexity. As such, there is likely an ideal way to integrate both human and AI teammates to create a collaborative training program. As AI teammates become more prevalent, this type of training will likely grow in utility and should, in turn, grow in priority.

As such, two actions are likely needed in the HAT research community moving forward. First, at a meta-level, the design and performance of human-AI teaming research studies should avoid having humans collaboratively trained prior to participating in a research study. If human-AI teaming researchers were to train humans prior to an experiment collaboratively, they may inadvertently worsen the performance of humans and teams, so this insight should be considered during the design of experimental procedures. Second, additional efforts are needed to explore the intelligent design of task-focused training in human-AI teams. The takeaway from this work should not be that collaborative training is bad for human-AI teams; rather, the existing design of collaborative training likely needs to change so that it can be effective for human-AI teams. As such, future research efforts should explore different types of collaborative training that use a mixture of human and AI teammates. In turn, this better collaborative training will help humans cultivate the needed skills to perform and learn in human-AI teams without requiring significant adjustment on their part.

## Conclusion

With an increasing prevalence of human-AI teaming research, the inevitable integration of human-AI teams in the real world continues to approach. Currently, humans are not prepared to collaborate and team with AI teammates, and

direct intervention will be needed to help them develop the skills and understanding needed to work with AI. As such, the human-AI teaming domain needs to continue to prioritize the creation and validation of novel training techniques to prepare humans. This paper presents a first-of-its-kind empirical research study that explores the design of an empirical experiment that explores the design and effectiveness of task-focused training in humans. Through this experiment, results find that having humans learn task-related skills collaboratively can negatively impact their performance when they eventually transition into a human-AI team. With this insight, human-AI teaming researchers and practitioners need to caution against collaboratively training humans prior to working in a human-AI team. Further, future research and application efforts should iterate on this work to create task-focused training methods that are especially beneficial to human-AI teams.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iDs

Christopher Flathmann  <https://orcid.org/0000-0002-5448-2610>  
 Beau G. Schelble  <https://orcid.org/0000-0003-3704-697X>

### References

- Bansal, G., Nushi, B., Kamar, E., Weld, D. S., Lasecki, W. S., & Horvitz, E. (2019, July). Updates in human-ai teams: Understanding and addressing the performance/compatibility tradeoff. *Proceedings of the AAAI conference on artificial intelligence*, 33(1), 2429–2437.
- Chen, J. Y., Lakhmani, S. G., Stowers, K., Selkowitz, A. R., Wright, J. L., & Barnes, M. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical Issues in Ergonomics Science*, 19(3), 259–282.
- Ellington, J. K., & Dierdorff, E. C. (2014). Individual learning in team training: Self-regulation and team context effects. *Small Group Research*, 45(1), 37–67.
- Endsley, M. R. (2023). Supporting human-AI teams: Transparency, explainability, and situation awareness. *Computers in Human Behavior*, 140, 107574.
- Entin, E. E., & Serfaty, D. (1999). Adaptive team coordination. *Human Factors*, 41(2), 312–325.
- Farrow, E. (2022). Determining the human to AI workforce ratio—exploring future organisational scenarios and the implications for anticipatory workforce planning. *Technology in Society*, 68, 101879.
- Flathmann, C., Schelble, B. G., Rosopa, P. J., McNeese, N. J., Mallick, R., & Madathil, K. C. (2023). Examining the impact of varying levels of AI teammate influence on human-AI teams. *International Journal of Human-Computer Studies*, 177, 103061.
- Flathmann, C., Duan, W., Mcneese, N. J., Hauptman, A., & Zhang, R. (2024). Empirically understanding the potential impacts and process of social influence in human-AI teams. *Proceedings of the ACM on Human-Computer Interaction*, 8, 1–32.
- Gorman, J. C., Cooke, N. J., Winner, J. L., Duran, J. L., Pedersen, H. K., & Taylor, A. R. (2007). Knowledge training versus process training: The effects of training protocol on team coordination and performance. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 51, No. 4, pp. 382–386). SAGE Publications.
- Ji, J., Qiu, T., Chen, B., Zhang, B., Lou, H., Wang, K., & Gao, W. (2023). Ai alignment: A comprehensive survey. *arXiv preprint arXiv:2310.19852*.
- Johnson, C. J., Demir, M., McNeese, N. J., Gorman, J. C., Wolff, A. T., & Cooke, N. J. (2023). The impact of training on human-autonomy team communications and trust calibration. *Human Factors*, 65(7), 1554–1570.
- Liang, D. W., Moreland, R., & Argote, L. (1995). Group versus individual training and group performance: The mediating role of transactive memory. *Personality and Social Psychology Bulletin*, 21(4), 384–393.
- Lyons, J. B., aldin Hamdan, I., & Vo, T. Q. (2023). Explanations and trust: What happens to trust when a robot partner does something unexpected? *Computers in Human Behavior*, 138, 107473.
- McNeese, N. J., Demir, M., Cooke, N. J., & Myers, C. (2018). Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human Factors*, 60(2), 262–273.
- McNeese, N. J., Flathmann, C., O'Neill, T. A., & Salas, E. (2023). Stepping out of the shadow of human-human teaming: Crafting a unique identity for human-autonomy teams. *Computers in Human Behavior*, 148, 107874.
- Nikolaïdis, S., Lasota, P., Ramakrishnan, R., & Shah, J. (2015). Improved human-robot team performance through cross-training, an approach inspired by human team training practices. *The International Journal of Robotics Research*, 34(14), 1711–1730.
- Marks, M. A., Sabella, M. J., Burke, C. S., & Zaccaro, S. J. (2002). The impact of cross-training on team effectiveness. *Journal of Applied Psychology*, 87(1), 3.
- O'Neill, T., McNeese, N., Barron, A., & Schelble, B. (2022). Human-autonomy teaming: A review and analysis of the empirical literature. *Human factors*, 64(5), 904–938.
- Oswald, F. L., Endsley, M. R., Chen, J., Chiou, E. K., Draper, M. H., McNeese, N. J., & Roth, E. M. (2022, September). The national academies board on human-systems integration (BOHSI) panel: Human-AI teaming: research frontiers. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 66, No. 1, pp. 130–134). SAGE Publications.
- Prichard, J. S., Stratford, R. J., & Bizo, L. A. (2006). Team-skills training enhances collaborative learning. *Learning and Instruction*, 16(3), 256–265.
- Prichard, J. S., & Ashleigh, M. J. (2007). The effects of team-skills training on transactive memory and performance. *Small Group Research*, 38(6), 696–726.
- Salas, E., DiazGranados, D., Weaver, S. J., & King, H. (2008a). Does team training work? Principles for health care. *Academic Emergency Medicine*, 15(11), 1002–1009.
- Salas, E., DiazGranados, D., Klein, C., Burke, C. S., Stagl, K. C., Goodwin, G. F., & Halpin, S. M. (2008b). Does team training improve team performance? A meta-analysis. *Human Factors*, 50(6), 903–933.

- Sawyer, T., & Gray, M. M. (2016). Procedural training and assessment of competency utilizing simulation. In *Seminars in perinatology* (Vol. 40, No. 7, pp. 438–446). WB Saunders.
- Schelble, B. G., Flathmann, C., McNeese, N. J., Freeman, G., & Mallick, R. (2022). Let's think together! Assessing shared mental models, performance, and trust in human-agent teams. *Proceedings of the ACM on Human-Computer Interaction*, 6, 1–29.
- Shen, H., Knearem, T., Ghosh, R., Alkiek, K., Krishna, K., Liu, Y., & Jurgens, D. (2024). Towards bidirectional human-AI alignment: A systematic review for clarifications, framework, and future directions. *arXiv preprint arXiv:2406.09264*.
- Verhoeven, Y., & Preuss, M. (2020). On the potential of rocket league for driving team AI development. In *2020 IEEE symposium series on computational intelligence (SSCI)* (pp. 2335–2342). IEEE.
- Zhang, R., McNeese, N. J., Freeman, G., & Musick, G. (2021). “An ideal human” expectations of AI teammates in human-AI teaming. *Proceedings of the ACM on Human-Computer Interaction*, 4, 1–25.